

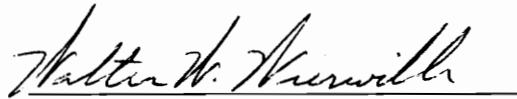
AN ACCURACY ANALYSIS OF TECHNIQUES FOR MEASURING
THE DURATIONS OF IN-CAR MANUAL TASKS

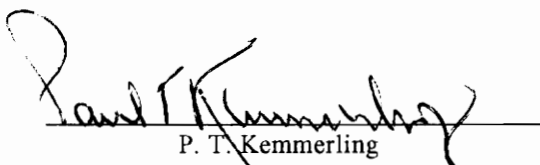
by

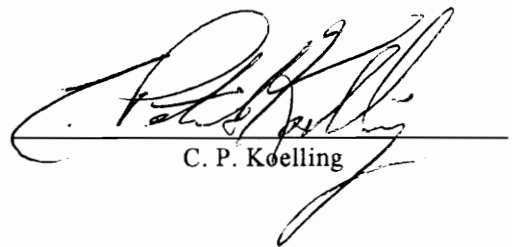
Barry Steven Grant

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE
in
Industrial and Systems Engineering

APPROVED:


W. W. Wierwille, Chairman


P. T. Kemmerling


C. P. Koelling

May, 1992

Blacksburg, Virginia

2

LD
5655
V855
1992
G726
c. 2

2

AN ACCURACY ANALYSIS OF TECHNIQUES FOR MEASURING
THE DURATIONS OF IN-CAR MANUAL TASKS

by

Barry Steven Grant

Committee Chairman: Dr. Walter W. Wierwille

Industrial and Systems Engineering

(ABSTRACT)

Human factors research of automobile driver behavior often calls for timing in-car manual tasks. By measuring the time that a driver directs resources away from the primary task of driving, investigators can determine both if and where improvements need to be made in the driver-vehicle relationship. The present study was designed to compare the accuracy of various techniques for measuring in-car manual task durations. Additionally, this research was intended to reveal how closely these techniques approach the accuracy of the frame-by-frame video analysis method, which is time-consuming and expensive to perform.

The study required three male and three female subjects to use an electronic stopwatch to measure "hand-off-wheel" times for 30 driver tasks. Each subject performed this procedure three times: once while sitting as an observer in the back seat of a research vehicle, once while watching a real-time video recording of task performance, and once while watching a one-sixth real-time video recording of task performance.

Three independent variables were examined: timing Method (three levels), Duration of in-car task (three levels), and Subject (six levels). Dependent measures gathered were raw timing error (a measure of response *bias*), absolute timing error (a measure of response *accuracy*), and squared timing error (a measure of response *consistency*). Timing error

was obtained by subtracting the measured time for a particular task from the “true” task time obtained by using the frame-by-frame video analysis technique.

Analysis of the data indicated a significant effect of Method on response bias. Specifically, use of the slow-motion video technique resulted in overestimation of in-car task durations, and use of the two real-time techniques resulted in estimates of task durations that were either equal to or less than the true durations. Significant effects of Subject, Gender, and Subject x Method were also revealed. The results suggest that the on-road timing technique should be used in the future, since this procedure requires little in terms of cost and implementation time, and errors are small when compared with the frame-by-frame technique. Furthermore, it is recommended that only one experimenter serve as a timing analyst during an experiment to ensure a constant level of accuracy.

To my family: Jerry, Sheila, and Bonnie

ACKNOWLEDGEMENTS

This project was possible due to the financial support of General Motors' Systems Engineering division. In particular, appreciation is extended to Dr. Brian Repa, Dr. Linda Angell, and Dr. Raymond Kiefer for their participation in the development and the progress of this study.

I would also like to express my gratitude to Dr. Walter Wierwille, who provided me with both the opportunity to undertake this research and the motivation to continually work to the best of my ability. Without Dr. Wierwille's advice, expert insight, and technical knowledge, this project would not have been possible. My thanks are also extended to Professor Paul Kemmerling and Dr. Pat Koelling for their valuable input into this study. Additionally, I would like to thank Dr. Robert Williges, Dr. Harry Snyder, and Dr. Jeff Woldstad for their most beneficial advice on experimental design and statistics.

Appreciation goes out to Mimi Buchanan for her assistance in planning this project and gathering the data, and to Michael Mitchell for his help in analyzing the data. Special thanks to Jean Harris, also, for taking the time to proofread my masterpiece.

Finally, I must thank my family, Jerry, Sheila, and Bonnie, for supporting me throughout this entire effort. Because of their unending love and their continual encouragement, I have been able to achieve yet another goal which I have set for myself. This thesis is dedicated to them.

TABLE OF CONTENTS

	<u>Page</u>
ABSTRACT	ii
DEDICATION.....	iv
ACKNOWLEDGEMENTS.....	v
LIST OF TABLES	ix
LIST OF FIGURES	x
INTRODUCTION	1
Background	1
Research Objectives	3
LITERATURE REVIEW.....	4
Overview	4
Human Servo Technique	4
General rules of simple reaction time.....	5
General rules of movement time	21
Studies using a human servo technique	23
General conclusions regarding the human servo technique.....	26
Frame-By-Frame Technique	27
Potential problems	27
General conclusions regarding the frame-by-frame technique.....	32
Alternate Task Duration Measurement Techniques	33
Summary	34

	<u>Page</u>
METHOD	36
Subjects.....	36
Apparatus.....	36
Experimental Design	44
Procedure.....	46
Experimental tasks.....	46
Experimental protocol	53
Research Hypotheses	54
RESULTS.....	58
Analysis of Raw Error Scores	60
Analyses of Absolute Error Scores.....	70
Analysis of Squared Error Scores.....	70
Analyses of Winsorized Data.....	75
Determination of Limits for Subject Responses.....	79
Time Requirements for Manual Timing Techniques.....	82
DISCUSSION.....	85
Interpretation of Results	85
Difficulties with Experimental Apparatus and Procedures.....	92
Conclusions	95
Suggestions for Future Research	97
REFERENCES.....	99

	<u>Page</u>
APPENDICES.....	104
A. Task Duration and Error Log Sheets	104
B. In-Car Manual Tasks	109
C. General Instructions for the Experiment	113
D. Participant’s Informed Consent Form.....	117
E. Analyses of Winsorized Data.....	120
VITA.....	125

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1. Counterbalancing of human servo Method presentation order	47
2. Construction of Quasi- <i>F</i> ratios and calculation of their tabled degrees of freedom	61
3. ANOVA summary table for raw error scores	62
4. Significant results of unconfounded pairwise comparisons for the raw-error-scores Subject x Method interaction, using the Tukey Honestly Significant Difference test.....	66
5. Results of chi-square analyses of Task means	68
6. Results of binomial tests of Task means (control combinations)	71
7. Results of Student <i>t</i> -tests (all three dependent variables)	72
8. ANOVA summary table for absolute error scores	73
9. ANOVA summary table for squared error scores.....	76
10. Significant results of unconfounded pairwise comparisons for the absolute-error-scores Subject x Method interaction (Winsorized data), using the Tukey Honestly Significant Difference test	81
11. Means, standard deviations, 10th percentile raw error scores, and 90th percentile raw error scores for four different task types	83
 E.1. ANOVA summary table for Winsorized raw error scores	 121
E.2. ANOVA summary table for Winsorized absolute error scores.....	122
E.3. ANOVA summary table for Winsorized squared error scores	123
E.4. Results of Student <i>t</i> -tests for Winsorized data	124

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. Mean reaction time as a function of constant foreperiod duration.....	13
2. Mean reaction time as a function of variable foreperiod duration	14
3. Probability density function of the random timing error (X) occurring at the beginning of the behavior interval	29
4. Probability density function of the random timing error (Y) occurring at the end of the behavior interval.....	30
5. General probability density function of the random timing error (u) due to sampling.	31
6. Research vehicle used in this study -- 1989 Oldsmobile Cutlass Supreme.....	37
7. Cronus 4 LED electronic stopwatch.....	40
8. Video analysis station	41
9. Video quad image viewed while measuring in-car manual task durations with the video servo techniques and the frame-by-frame technique	42
10. Close-up view of VCR remote control used during video analysis segments of the study	43
11. Experimental cell design	45
12. Critical path charts for scheduling subjects.....	48
13. Mean timing error as a function of human servo Method.....	64
14. Subject x Method interaction (raw error data).....	65
15. Mean timing error as a function of Task	67
16. Mean absolute timing error as a function of Subject.....	74
17. Scatterplot of the 540 raw error scores.....	77
18. Winsorized scatterplot of the 540 raw error scores.....	78
19. Subject x Method interaction (Winsorized absolute error data)	80

INTRODUCTION

Background

The human factors research literature contains numerous reports of studies in which behavior duration serves as a dependent variable. Behavior duration refers to the amount of time involved from the onset of an initiating stimulus until the corresponding human response is completed (Meister, 1985). This duration can be extremely short, as in the case of a simple reaction-response paradigm, or it can be more lengthy, as in the time required to perform a complicated task.

Although there are many situations in which the measurement of behavior durations can be used to provide insight into human performance, one that is of particular interest is that where a driver is performing in-car manual tasks while driving (i.e., tuning the radio, adjusting the fan speed, etc.). When an automobile is travelling at high speeds, optimal driver performance is essential. If the driver's attention or behavior is diverted away from the driving task for too long, there is an increased potential for an accident. By measuring the time that a driver directs resources away from the primary task of driving, researchers can determine both if and where improvements need to be made in the driver-vehicle relationship.

Several fairly objective techniques have been used in transportation research to assess the amount of time that it takes drivers to perform in-car manual tasks. These methods typically involve the use of some sort of timing device. One popular way of measuring in-car manual task durations is with a manually operated mechanical, electronic, or electro-mechanical timer. With this technique, an observer holds some type of control (often a pushbutton) that is attached to a timing device. When the driver initiates the task of interest, the observer activates the control that begins the timer. Likewise, the observer deactivates the control when the task is concluded, and the timer stops running. The time

elapsed from the beginning of the driver's response until the end of that response is then recorded by the observer, and the timer is reset. This technique is usually called the "human servo" technique.

A more objective way of measuring in-car manual task durations is with videography. This system is used for recording the desired behaviors onto videotape. By using a time code generator to imprint a time stamp on each frame of the recording, an observer can later review each in-car task frame-by-frame to determine exact start and stop points. The difference between the start time and the stop time for a particular task equals the task duration. This technique is referred to as the "frame-by-frame" technique.

Because the measurement of in-car manual task durations is so important in transportation research, it is necessary to examine the accuracy, bias, and consistency of the various timing procedures which are used to make these measurements. Although the frame-by-frame technique can be considered to be practically free of significant timing errors, there are costs associated with this preciseness: the high price of equipment and the amount of time required for data analysis. The human servo technique, on the other hand, although requiring only a minimal investment in equipment and a short amount of time for data analysis, may suffer from accuracy, bias, and consistency problems resulting from simple reaction/response delays of the observer. If research could demonstrate that some variation of the human servo technique allows for preciseness approximating that of the frame-by-frame technique, it might be possible to greatly simplify the measurement of both in-car manual task durations and other behavior durations as well.

Research Objectives

The primary research objectives of this study are as follows:

Objective 1: To compare the accuracy, bias, and consistency of three variations of the human servo technique, and to determine if any of these variations approach the preciseness of the frame-by-frame method of measuring in-car manual task durations.

Objective 2: To determine the effects of in-car manual task length on the accuracy, bias, and consistency of task duration measurements made using the human servo technique.

Objective 3: To examine the effects of individual subject differences on the accuracy, bias, and consistency of task duration measurements made using the human servo technique.

LITERATURE REVIEW

Overview

This literature review begins with a discussion of factors which affect the human servo technique for measuring in-car manual task durations. These factors can be generally classified as either simple reaction time factors or movement time factors. Next, several studies in which a variation of the human servo technique has been used will be discussed. The frame-by-frame technique will then be compared and contrasted with the human servo procedure. The literature review will conclude with a discussion of alternate techniques for measuring in-car manual task durations. The intent of this discourse is to provide insight into those aspects of *timing analyst (observer)* performance which may contribute to errors in the recording and/or analysis of in-car manual task duration data.

Human Servo Technique

To discuss the human servo technique in detail, it is first necessary to describe a common scenario within which this procedure is used. In this scenario there is an experimenter, a timing analyst (observer) and a subject (driver) riding in an automobile. The experimenter, in the front passenger seat, is responsible for instructing the driver to perform an in-car manual task. The observer, situated in the back seat of the automobile, is responsible for timing the driver's performance of that task.

Assuming that the back-seat observer is recording the amount of time that the driver's hand is off the steering wheel (hand-off-wheel time is a commonly used measure of in-car manual task duration), that observer presses a pushbutton control connected to an electronic timer at the instant that the driver's hand comes off the wheel. This action initiates timing of the driver's behavior. The observer closely monitors the driver's performance of the task until the hand returns to the steering wheel. At the instant the driver's hand touches

the wheel, the observer once again presses the pushbutton control. The elapsed time to perform the task is displayed on the timer, and it is recorded by the observer before the timer is reset for the next trial.

By breaking this scenario into small components, it becomes obvious where there may be errors in data collection and analysis. Due to a simple reaction time lag, there will usually be a slight delay between the moment the driver takes a hand off the steering wheel and the moment the observer reacts/responds to this action by pressing the pushbutton. Similarly, there will often be a simple reaction time lag between the time when the driver's hand makes return contact with the steering wheel and the time at which the observer responds. If the beginning lag and the ending lag are exactly equal in length, the obtained time reading will be accurate. If, however, one lag is shorter or longer than the other, inaccuracy is introduced. To better understand the issues at hand, an understanding of simple reaction time and the factors affecting it is required.

General rules of simple reaction time. Reaction time is typically defined as the time from the beginning of an unambiguous stimulus until the beginning of a response to that stimulus (Boff and Lincoln, 1988; Huchingson, 1981; Sanders and McCormick, 1987). It should be noted that this time does not include the actual movement time in responding to the stimulus; it only includes the time until the movement is initiated (Sanders and McCormick, 1987). "Response time," on the other hand, refers to the combination of reaction time and movement time (Huchingson, 1981).

There are actually several different types of reaction time, depending upon how many distinct stimuli and responses are involved. In examining the human servo technique, however, simple reaction time demands principal consideration. In a simple reaction time task there is just one stimulus and the subject can make only one response to that stimulus

(Boff and Lincoln, 1988). Because of its inherent simplicity, simple reaction time is typically the smallest of all reaction times (Sanders and McCormick, 1987).

How long are the fastest simple reaction times? Of course, the answer varies depending upon the factors involved. In general, however, most authors agree that a simple reaction to a strong, anticipated signal takes from about 100 to 200 milliseconds (Grandjean, 1986; Huchingson, 1981; Keele, 1986; Sanders and McCormick, 1987; Van Cott and Kinkade, 1972). In a related study Wargo (1967) broke simple reaction time into its physiological components to arrive at an estimate of 113 to 328 milliseconds to react to a single stimulus. These physiological components of simple reaction time are receptor delays, afferent transmission delays, central process delays, efferent transmission delays, and muscle latency and activation time. While this literature review will not dwell upon the fine psychophysiological details of simple reaction time, it is important to note that the whole process is rather complicated.

As mentioned above, 100 to 200 milliseconds is the accepted estimate of simple reaction time in an ideal situation. In many situations conditions are far from ideal, however, and a host of factors may increase that 100 to 200 milliseconds estimate. For example, Boff and Lincoln (1988) list sensory modality, location of stimulus in field of view, foreperiod attributes, perceptual characteristics of the stimulus, mode of response, motivation, fatigue, and response device as factors which can play a significant role in determining a subject's simple reaction time to a stimulus. Sanders and McCormick (1987) add stimulus discriminability, compatibility of stimulus and response, repetition, and required accuracy to Boff and Lincoln's (1988) list. Many other authors discuss these same factors in considerable detail (Huchingson, 1981; Johnson, 1923; Teichner, 1954; and Wickens, 1984).

In analyzing the human servo procedure scenario established earlier, most of the factors mentioned in the previous paragraph do not come into play. There is only one stimulus (driver's hand leaving the steering wheel or returning to it), the observer is usually looking directly at the stimulus, there is a rest period between each trial (reducing the effects of vigilance issues), and the pushbutton response is itself simple. Only two factors appear to have the potential for affecting observer simple reaction time in using the human servo procedure. These factors are age and foreperiod attributes.

In a study by Wilkinson and Allison (1989), 5325 men, women, and children attending a public exhibition performed a one-minute test of simple reaction time. Reaction time was averaged across the last eight of ten trials presented to the participants. The results showed that average simple reaction time gradually increases from the twenties up to age sixty and over. Furthermore, simple reaction time increases quite rapidly up to the twenties.

Another study of the effects of age on simple reaction time revealed similar results. Gottsdanker (1982) asked 220 subjects ranging from 18 to 93 years of age to respond to tones as quickly as possible by pressing a pushbutton. He, like Wilkinson and Allison (1989), found that simple reaction time is a gradually increasing function of age. This increase, although statistically significant, was only about two milliseconds per decade.

These results have definite implications for using the human servo technique to measure in-car manual task durations. Apparently, the age of the timing analyst may influence reaction time in starting and stopping the timer. Judging from Wilkinson and Allison's (1989) and Gottsdanker's (1982) studies, older observers will be a little slower in responding to the driver's hand movements off or onto the steering wheel than will younger observers. However, considering the amount of time that a driver is usually engaged in a secondary task, these simple reaction time lags are probably too small to affect

the duration measurements to any significant degree. Furthermore, if the simple reaction time delay at the beginning of the timing interval is equal to the delay at the end of the interval, duration measurement accuracy will not be affected.

According to Niemi and Näätänen (1981), a typical simple reaction time paradigm includes five main factors: the warning signal, the foreperiod (also called the warning interval or preparatory interval), the reaction stimulus, the response, and the intertrial interval. Of these five factors, the foreperiod appears to have the greatest potential for affecting the simple reaction time of an observer using the human servo procedure. As such, a description of this foreperiod is in order. Niemi and Näätänen (1981) explain that in a normal simple reaction situation the warning signal precedes the reaction stimulus and is intended to inform the subject of the approaching stimulus. The empty period between the warning signal and the reaction stimulus is called the foreperiod. This foreperiod provides what Niemi and Näätänen (1981) call a “temporal frame of reference” during which a subject prepares to respond to the reaction stimulus. If the foreperiod length is estimated correctly, a subject can time the onset of his or her reaction to occur almost simultaneously with the onset of the reaction stimulus.

If one carefully considers the predefined scenario for using the human servo technique, it becomes obvious why the foreperiod may play such a vital role in influencing the observer’s simple reaction time. In the case where the driver *begins* to perform an in-car manual task, the warning signal is the start of the verbal instructions given by the front-seat experimenter to the driver. The reaction stimulus is the driver’s hand leaving the steering wheel. Consequently, the foreperiod is the time between the giving of the instructions and the hand leaving the wheel. Finally, the response is the observer’s pressing of the timer pushbutton. The intertrial interval is not directly applicable to this scenario.

As for the case where the driver *finishes* performing the in-car manual task, the warning signal is the beginning of the hand movement back toward the wheel and the reaction stimulus is the hand making return contact with the wheel. The foreperiod in this second situation is the time between the hand beginning to return to the steering wheel and the hand actually touching the wheel. Lastly, the response is the observer's pressing of the timer pushbutton.

Because the foreperiods of concern in this scenario are destined to be variable and somewhat unpredictable, a question arises as to how the observer's simple reaction times will be affected. A host of studies on foreperiod characteristics and their relation to simple reaction time should help to clarify how timing analysts can be expected to behave when using the human servo technique.

One of the first questions ever raised about simple reaction time foreperiods was whether there is an optimal duration (Niemi and Näätänen, 1981). In a landmark article Woodrow (1914) stated that a two-second foreperiod is optimal if foreperiod duration is to remain constant. He continued, however, by saying that a twelve-second foreperiod is optimal if the length of the preparatory interval is varied irregularly. Regardless of foreperiod constancy or variability, Woodrow found that simple reaction time increases as the warning interval duration deviates from the optimum.

A study by Telford (1931) using auditory stimuli revealed somewhat different results from those obtained by Woodrow (1914). Telford found that a one-second foreperiod is optimal when foreperiod durations are variable, as opposed to constant. This differs from Woodrow's finding of a 12-second optimal foreperiod duration under variable conditions. Telford, however, was not the first to disagree with Woodrow's optimal foreperiod length conclusions. In a study of the effects of foreperiod duration and variability on simple reaction time, Karlin (1959) found that reaction time to a 240-Hertz, 33 db tone rises

without inflection as constant foreperiod duration is increased from 0.5 to 3.5 seconds. In two other studies, Bertelson (1967) and Bertelson and Tisseyre (1967) found that the fastest reaction times in a choice reaction time task correspond to foreperiod durations ranging from 100 to 150 milliseconds. In the first study (Bertelson, 1967), short constant foreperiod durations were used and subjects were required to respond to either of two small lamps after hearing a warning click. In the second study (Bertelson and Tisseyre, 1967), short variable foreperiod durations were used and everything else remained the same as in the first experiment. Regardless of foreperiod variability, they found the optimal foreperiod duration to be quite small.

Perhaps the conflicting results cited above can be partially explained through a study performed by Okamoto (1980). He tested six subjects' simple reaction times to an LED stimulus under variable foreperiod conditions. Under the short foreperiod conditions, subjects responded to stimuli following warning intervals of 1.00, 1.30, 1.69, 2.19, 2.84, and 3.69 seconds. Under the long foreperiod conditions, subjects responded to stimuli following warning intervals of 2.84, 3.40, 4.07, 4.88, 5.85, and 7.01 seconds. Okamoto found that the optimal foreperiod duration under the short foreperiod conditions was between 2.19 and 2.84 seconds, and the optimal foreperiod duration under the long foreperiod conditions was between 4.88 and 5.85 seconds. That is, the optimal foreperiod length in any given set of trials depends upon the range from which the foreperiods are sampled. This finding might explain why numerous scientific studies of optimal foreperiod duration disagree with each other. The differences may all be an artifact of foreperiod length range. Teichner's (1954) statement that ". . . a great many factors influence the optimum foreperiod" is indeed true.

One of the factors affecting the optimal foreperiod duration for a given simple reaction is stimulus expectancy. Expectancy, as it is used here, is defined as the subjective

probability of the immediate occurrence of the reaction stimulus (Näätänen, 1970). According to Näätänen, Muranen, and Merisalo (1974), expectancy is apparently related to reaction time in such a way that the degree of expectancy plays a major role in determining how prepared a subject is to respond to a stimulus. Näätänen et al. (1974) attempted to determine the moment of greatest expectation in a simple reaction time experiment. Four subjects were asked to predict the moment of occurrence of a stimulus by pressing a key, under constant preparatory intervals of 0.25, 0.5, 1, 2, and 4 seconds. Results indicated that after 0.5 second both accuracy of prediction and simple reaction speed decrease continuously. Thus, Näätänen et al. would recommend an optimal foreperiod length of 0.5 second under constant foreperiod conditions. This recommendation, of course, differs from the recommendations of Woodrow (1914), Bertelson (1967), and Bertelson and Tisseyre (1967). It is hypothesized that either foreperiod range or some other unaccountable variable caused this discrepancy.

Another important foreperiod-related factor that can affect simple reaction time is uncertainty. According to Niemi and Näätänen (1981), an increase in foreperiod uncertainty leads to a decrease in maximum stimulus expectancy. This time-uncertainty is a function of both mean foreperiod duration and foreperiod variability (Klemmer, 1957). In a study by Klemmer (1957), both foreperiod duration and foreperiod variability were combined to derive a general measure of time uncertainty in bits. Klemmer found that simple reaction time to a neon stimulus bulb is a direct linear function of time uncertainty. That is, simple reaction time increases with an increase in foreperiod uncertainty (and decrease in stimulus expectancy).

Several of the aforementioned studies have alluded to the significance of constant versus variable warning intervals in the determination of the optimal foreperiod for a simple reaction. In the most fundamental foreperiod arrangement, foreperiod remains constant

throughout an experimental session and influences subjects' performance through their inability to estimate time perfectly (Niemi and Näätänen, 1981). In general, given constant foreperiod lengths, simple reaction time appears to increase as foreperiod duration increases, except in the case of very short warning intervals (Näätänen et al., 1974; Niemi and Näätänen, 1981; Teichner, 1954; Telford, 1931; Woodrow, 1914). Figure 1, adapted from Näätänen et al. (1974), illustrates an example of this effect.

It should be noted that reaction time will not rise indefinitely with a corresponding increase in constant foreperiod duration. In a 1965 study by Bevan, Hardesty, and Avant, six groups were exposed to simple visual signals, each group receiving a different constant foreperiod length (either 10, 20, 40, 80, 160, or 320 seconds). For each of these constant-interval groups, there was a corresponding variable-interval group that received foreperiod intervals of the the same *average* duration. For all groups, Bevan et al. found that simple reaction time increases with foreperiod duration up to a foreperiod length of 80 seconds. After that, reaction time appears to level off.

In addition to the results mentioned above, Bevan et al. (1965) found that constant foreperiod groups had significantly shorter reaction times than variable foreperiod groups. This phenomenon is generally the case when dealing with variable foreperiods, for individuals have little information to aid them in timing their responses to coincide with the onset of the reaction stimulus (Niemi and Näätänen, 1981). Furthermore, as first recognized by Woodrow (1914), simple reaction time tends to decrease with foreperiod length when variable foreperiod intervals are involved (Figure 2, adapted from Niemi, 1979, provides an example of this phenomenon). This effect is exactly the opposite of what is encountered when dealing with constant foreperiods. Several studies have confirmed this fact (Drazin, 1961; Mo and George, 1977; Nickerson, 1965; Polzella et al., 1989).

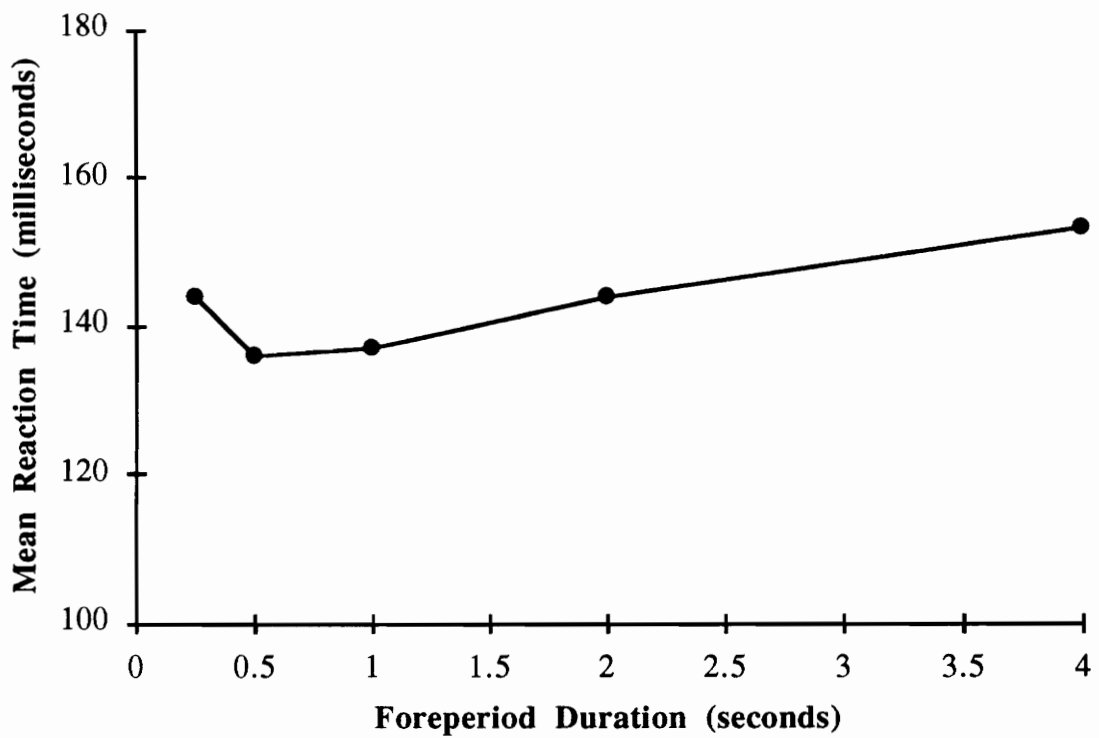


Figure 1: Mean reaction time as a function of constant foreperiod duration. Adapted from Näätänen et al., 1974.

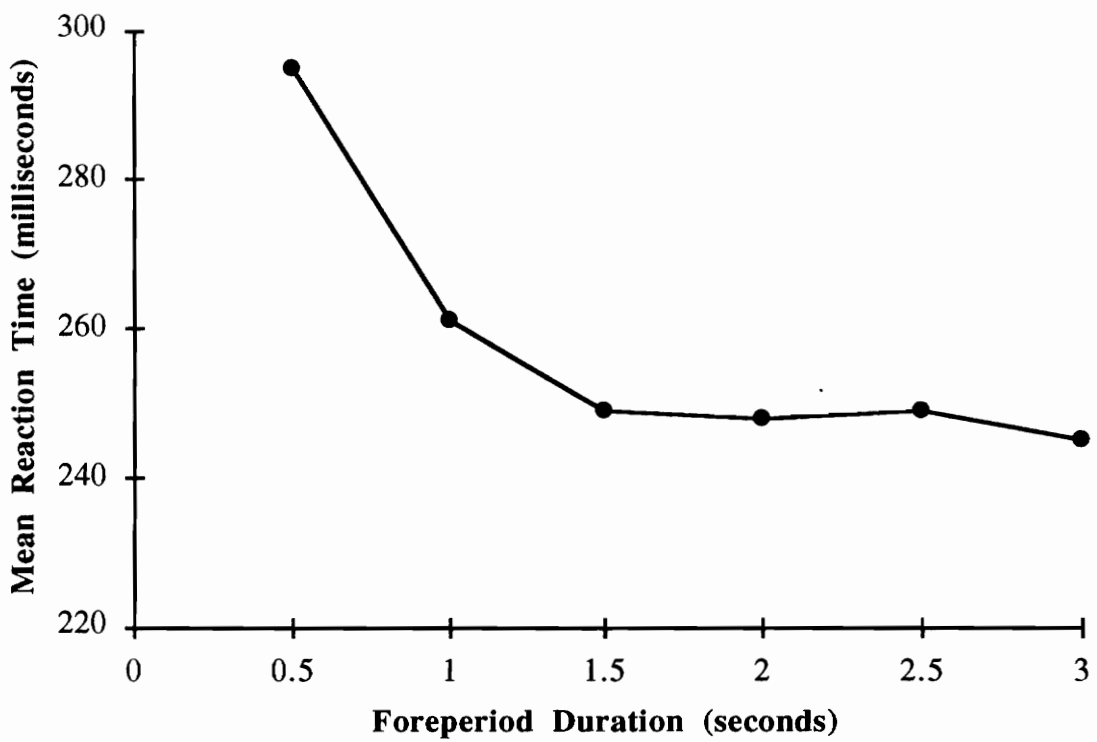


Figure 2: Mean reaction time as a function of variable foreperiod duration. Adapted from Niemi, 1979.

In a simple reaction time study by Drazin (1961), subjects were asked to respond to the onset of a lamp by pressing a key as quickly as possible. A tone preceded the reaction stimulus by variable durations; these durations were distributed rectangularly. In analyzing the results of the study, Drazin found that longer reaction times occur with shorter foreperiods. Furthermore, he found that the initial gradient of the reaction time/foreperiod curve becomes less steep as the minimum foreperiod increases. Wickens (1984) explained Drazin's results through the previously mentioned concept of expectancy. According to Wickens, the longer the time that has passed since the warning signal was given, the more the subject will expect the stimulus to occur. As a result, the criterion for response is lowered as time continues to pass, and the subject responds quickly when the stimulus finally occurs.

Mo and George (1977) and Polzella, Ramsey, and Bower (1989) offer the same sort of expectancy hypothesis to explain the phenomenon of decreasing reaction time with increasing variable foreperiod length. In the Polzella et al. study (1989), subjects reacted to a visual stimulus that was presented with a probability of either 1.00 or 0.90. Stimulus presentation followed foreperiods of 300, 400, 500, 600, or 700 milliseconds. They found that mean reaction time decreases linearly under 100% stimulus probability conditions and curvilinearly under 90% stimulus probability conditions as a function of increasing foreperiod duration.

In a study by Nickerson (1965), subjects were told to respond to the second of two visual signals presented to them on a CRT by pressing a telegraph key. These two signals differed by various durations. In analyzing the results, Nickerson found (like many others) that simple reaction time decreases with foreperiod length under variable foreperiod conditions. In a second study by Nickerson (1967), subjects also responded to a visual stimulus (a small lamp) following a warning interval of various durations. In this study,

however, subject expectancy was removed by using a computerized Bernoulli process to time the onset of the reaction stimuli. Interestingly, mean reaction time was found to increase with mean foreperiod duration. A follow-up study (Nickerson and Burnham, 1969) confirmed the results of the second Nickerson (1967) study. Apparently, expectancy does play a truly significant role in determining how a subject will react to variable foreperiods.

Buckolz and Wilberg (1975) discuss a second explanation for why simple reaction time decreases as warning interval length increases when variable foreperiods are involved. In addition to the expectancy hypothesis, they describe what they call a "previous foreperiod effect" to explain this phenomenon. The previous foreperiod effect assumes that a subject's expectancy of when a reaction stimulus will occur during the current foreperiod is influenced by the length of the previous foreperiod. This previous foreperiod effect occurs only when a long foreperiod is followed by a shorter foreperiod, the effect being that simple reaction time to the shorter foreperiod is increased in relation to the situation where sequential foreperiods are of the same length. When foreperiods are variable, the shorter ones are more often preceded by relatively longer foreperiods than are the longer ones. Thus, longer simple reaction times will tend to correspond to the shorter foreperiods.

Zahn and Rosenthal (1966) tested subjects' simple auditory reaction times in relation to both the length and the frequency of each member of two pairs (one and three seconds, and three and ten seconds) of foreperiods presented in a variable sequence. Their results confirmed that when a trial with a certain preparatory interval follows a trial with a longer foreperiod, the reaction time will be longer than if the preceding trial had a foreperiod of shorter or equal length. Thus, they helped to buttress the argument for a previous foreperiod effect.

Two other researchers found results opposing those of Zahn and Rosenthal (1966) and many others (Niemi and Näätänen, 1981). In the Drazin (1961) study described above, Drazin found sequential effects in both directions. However, in those situations where shorter foreperiods preceded longer foreperiods, there were larger sequential effects than in the opposite situation. Niemi and Näätänen (1981) propose that, as a result of Drazin's experimental design, subjects had no opportunity to acquire sufficient information to make predictions about stimulus onset time. A second study described earlier (Karlin, 1959) revealed similar results to those of Drazin (1961). Regardless of the support offered for Drazin's study by Karlin, however, the most salient feature of sequential effects is a lengthening of reaction time for a short foreperiod when it is immediately preceded by a longer foreperiod (Niemi and Näätänen, 1981).

Finally, Nickerson (1965) proposes a third explanation for why simple reaction time tends to increase as warning interval length decreases when variable foreperiods are involved. He quotes several researchers (Craik, 1948; Welford, 1952, 1959; Davis, 1956, 1957, 1959) who attribute the response delay corresponding to short warning intervals to the human's limitations as a signal processing system. According to Nickerson's discussion of the experimenters' theory, the human being has a "decision center" which can only handle one input at a given time. When two inputs arrive to this "center" close together in time, one must be stored until the other one has been adequately processed. Thus, if a warning signal were to precede a reaction stimulus by a very short amount of time, a subject may not be physically prepared to process the reaction stimulus right away. As a result, simple reaction time would increase.

A similar theory to that discussed by Nickerson (1965) is proposed by Possamaï, Granjon, Requin, and Reynard (1973) to explain the so-called previous foreperiod effect. In their study, Possamaï et al. analyzed the data from an earlier auditory reaction time study

(Requin, Granjon, Durup, and Reynard, 1972) in which three groups of ten subjects were exposed to different distributions of variable inter-stimulus intervals (analogous to warning intervals). They found that the shortest reaction times occurred when the inter-stimulus interval (ISI) for one trial equalled the inter-stimulus interval for the previous trial. Additionally, reaction time increased when the inter-stimulus interval for a previous trial was longer than that for the present trial. Possamaï et al. explain these results by stating that “the time-course of the preparatory process during ISI(n) is adjusted according to ISI(n-1) duration” (1973). What this means is that a subject will not be well-prepared to respond to a reaction stimulus following a short inter-stimulus interval if that short interval is preceded by a longer one. As a result, reaction time will suffer.

The next logical step would be to explain how all of the research on foreperiod length relates to observer simple reaction time when using the human servo technique to measure in-car manual task durations. First, the focus of this discussion will be upon the driver’s *commencement* of the in-car manual task. As previously mentioned, the beginning of the experimenter’s verbal instructions to the driver serves as the warning signal for the observer. Furthermore, the driver’s removal of the hand from the steering wheel serves as the reaction stimulus for the observer. The time interval between the start of the verbal instructions and the removal of the hand from the steering wheel is the foreperiod.

It can safely be said that the foreperiods occurring at the beginnings of tasks will not be of constant duration. They will vary with both length of verbal instruction and length of driver psychophysiological processing of the information presented. As a result, the scientific literature predicts that observer simple reaction time following the reaction stimulus will decrease with foreperiod length. This conclusion leads to the suggestion that verbal instructions by the experimenter be given slowly. Slow, clear instructions should

allow for the maximization of expectancy and the minimization of time uncertainty in the observer's preparation to respond.

The topic of sequential effects also bears mentioning. Based upon the work of several researchers (Buckolz and Wilberg, 1975; Niemi and Näätänen, 1981; Woodrow, 1914; Zahn and Rosenthal, 1966), it is predicted that the occurrence of relatively long beginning warning intervals before relatively short beginning warning intervals will lead to increased simple reaction times in the short beginning foreperiod trials. However, if the experimenter is careful to avoid excessively short instructions, this effect can be reduced.

In considering the driver's *completion* of the in-car manual task, one is faced with a different situation. In this case the warning signal is both the end of the driver's performance of the task and the beginning of the hand movement back toward the steering wheel. The return contact of the hand onto the wheel serves as the reaction stimulus for the observer. Finally, the time interval between the end of the task (beginning of the hand movement) and the contact with the steering wheel is the foreperiod.

In this situation the foreperiod length is again somewhat variable; it depends upon the distance of the task controls from the steering wheel and the driver's speed of movement. When observing the hand movement back to the wheel, however, the moment of occurrence of the reaction stimulus may be more predictable than it is for the first reaction stimulus.

Simon and Slaviero (1975) performed a simple reaction time study that deserves mentioning at this point. They had subjects respond to the onset of a stimulus light by pressing a key as soon as the light was activated. Half of the trials in this experiment were "countdown trials," whereby a sequence of lights provided information about the passage of time during a constant two-second foreperiod. The other half of the trials were merely regular trials with the same constant foreperiod length. Results revealed significantly faster

simple reaction times for the countdown trials than for the regular trials. According to Simon and Slaviero, the countdown procedure enabled the subjects to determine when the stimulus would occur and to establish an utmost state of readiness.

Similar results might be expected when using the human servo technique to measure in-car manual task durations. By watching the driver's hand move back toward the steering wheel, the timing analyst is provided with a type of countdown timer. As the hand gets closer and closer to the wheel, the probability of an instantaneous stimulus occurrence grows higher and higher. Time uncertainty is reduced, and expectancy is increased until the point where the hand makes contact with the wheel. Thus, it is possible that observer simple reaction time will be minimized when the pushbutton is pressed at the end of an in-car task.

There are several factors, however, which could prevent the observer's simple reaction time at the end of an in-car manual task from being faster than the simple reaction time at the start of a task. In fact, these factors could result in *slower* simple reaction times at the end of a task. First, there may be a point at which an observer cannot use a "countdown timer" to make his or her button-press coincide with the onset of the reaction stimulus. In other words, it appears possible that excessively short warning intervals will counteract the ability of an observer to predict with maximum certainty the occurrence of a reaction stimulus, even though some type of countdown timer is available. Next, it is conceivable that an observer might recognize a simple reaction lag at the beginning of a timing interval and compensate for this beginning lag by introducing an intentional lag at the end of the timing interval. If the observer could make the ending lag about equal in duration to the beginning lag, accuracy would not be sacrificed. Third, it is possible that the shorter foreperiod at the end of a task than at the beginning of a task could lead to a longer simple reaction at the end of the task. In fact, this possibility is supported by the

foreperiod research literature, which states that simple reaction time increases following short, variable foreperiods.

As mentioned at the beginning of this literature review, a timing analyst will obtain an accurate measurement of an in-car manual task duration if the beginning and ending simple reaction times are equal. If the beginning simple reaction time is shorter than the ending simple reaction time, the true task duration will be overestimated. Furthermore, If the beginning simple reaction time is longer than the ending simple reaction time, the true task duration will be underestimated. The present study was designed to reveal how, exactly, timing analysts respond when using the human servo technique under various conditions.

General rules of movement time. The topic of movement time, although not as critical in this review as the topic of simple reaction time, deserves mention. As stated earlier, response time is composed of both reaction time and movement time. Movement time is often defined as the time required to carry out a physical maneuver once a signal has been given. This time period begins when the reaction time ends (at the beginning of the physical maneuver), and it ends when the physical maneuver is completed (Huchingson, 1981).

According to Moore (1904), when a subject is told to react as quickly as possible, there is no relationship between simple reaction time and the ensuing movement time. O'Donnell (1972) also found no correlation between simple reaction time and movement time. However, Moore (1904) continues by stating that when a subject is told to both react *and move* as quickly as possible, the movement time remains constant and the reaction time tends to vary. He attributes this result to a subject's tendency to react in a "muscular" manner when instructed to make the quickest possible movement. Finally, Moore (1904) found that there is no effect of preparatory interval on movement time, although there is a

considerable effect on reaction time. This effect on reaction time is to be expected, based upon the findings of the foreperiod studies discussed earlier.

The work of Moore (1904) and O'Donnell (1972) should provide some insight into the effects of movement time on observer performance when using the human servo technique. Assuming that the observer is instructed to react and move as quickly as possible when activating and deactivating the timer, the movement times required in pressing the pushbutton should be relatively constant. Hence, low within-observer variance can be predicted. On the other hand, between-observer variance may be larger, given that humans differ in psychophysiological capabilities.

The question remains, however, of whether the movement time involved in pressing a pushbutton is large enough to result in errors in measuring in-car manual durations. According to Shinnick (1987), the MODAPTS (Modular Arrangement of Predetermined Time Standards) predetermined time standards dictate that the most time involved in pressing and releasing a pushbutton is 129 milliseconds. By dividing this time estimate in half, one can predict that the most time involved in simply pressing a pushbutton would be 65 milliseconds. On the other hand, Maynard's (1963) predetermined time standards (MTM -- Methods Time Measurement) predict a maximum button-press time of 72 milliseconds. Finally, in a study by Bradley (1975), subjects were required to release a fixed telegraph key (which started a timer) and press a one-inch diameter pushbutton (which stopped the timer) as quickly as possible. The average time to perform this task, disregarding any errors in pressing the pushbutton, was between 100 and 150 milliseconds. This last time range provides an *extremely* conservative estimate of how long it takes to press a pushbutton, for the subjects' fingers were not on the pushbutton from the beginning.

Given the aforementioned movement durations, which are quite small, and provided that the pushbutton control used by the observer activates the timer after only a slight downward movement, it can be assumed that the movement time in operating the timer has a negligible effect on the measurement of in-car manual task durations. Compared to the length of time involved in recording an in-car task, the time involved in pressing a pushbutton can be presumed to be inconsequential. Furthermore, if one can assume that movement time in pressing the pushbutton at the beginning of a task is equal to movement time in pressing the pushbutton at the end of a task, the button-press can be almost guaranteed to have no effect on timing accuracy.

Studies using a human servo technique. Several studies report using a variation of the human servo technique for recording behavior durations. One of the most insightful experiments was performed by Johansson and Rumar (1971). The objectives of their investigation were twofold: (1) they hoped to obtain a correction factor for estimating automobile driver brake reaction time when braking was unexpected from brake reaction time measures obtained when braking was expected, and (2) they wanted to determine the distribution of brake reaction times in a representative sample of drivers under unexpected braking conditions. As part of the experimental procedure, every car heading southbound on a particular road was stopped by the police and the driver was asked to participate in an experiment. If the driver agreed, he or she was told that a loud Klaxon horn would be heard next to his or her car sometime over the next ten kilometers, and this would be the signal to begin braking. About five kilometers from the police barrier was a concealed horn and measuring station. When the car passed this point, a photocell triggered the horn and an electronic timer was activated. A well-hidden assistant on the side of the road simultaneously watched for the car's brake lights to glow. As soon as he saw this, he pressed a contact key which stopped both the horn and the electronic timer.

This experiment provides a perfect example of the human servo technique in practice. Although the actual outcome of the experiment is not important to this review (see Johansson and Rumar, 1971 for more details), the information obtained on experimenter accuracy is. Johansson and Rumar explained that the accuracy of the measurements was dependent on the reliability of the assistant's reaction time. As such, they maintained, the standard error of his measurement time should have been negligible compared to that of the brake reaction time. To this end, they set a criterion stating that the variance of the assistant's reaction time should be no more than 20% of the variance of the brake reaction time.

Johansson and Rumar controlled the assistant's accuracy in two ways. First of all, they chose an assistant who demonstrated a relatively constant simple reaction time in pretrials. He was then trained to respond in as constant a manner as possible. Second, they "calibrated" the assistant in relation to his reaction time 322 times during the course of the study. Judging by the range of calibration measurements, they determined that the assistant's reaction time spread could not shift any single brake reaction time by more than 0.05 second in either direction.

A social psychology study by Goldberg and Wellens (1979) also involved the utilization of a human servo technique. This study was designed to compare direct face-to-face interviews with television-mediated interviews in regard to a hypothesized compensatory relationship between intimacy level of conversation topic and subjects' looking/smiling behaviors. Within both the face-to-face interview situation and the television-mediated interview situation, subjects' glances, smiling, talking, and listening behaviors were recorded on videotape. At the end of the study, four raters simultaneously viewed the videotaped interviews from four separate cubicles. Each rater pressed a hand-

held pushbutton (connected to an Esterline Angus event recorder) for the duration of time the behavior they were judging occurred.

After all ratings were finished, the raters rotated behaviors they were rating and a second set of ratings was made so as to determine interrater reliability. Pearson product-moment correlations were calculated between the two sets of ratings and the results were as follows: looking time, 0.94; smiling time, 0.90; talking time, 0.93; listening time, 0.88. Since these reliability coefficients were so high, all later analyses were based upon the average of the two sets of ratings.

Similar use of a human servo technique can be found in other social psychology studies. For example, Cherulnik, Neely, Flanagan, and Zachau (1978) were interested in testing the hypothesis that individuals with well-developed social skills will demonstrate higher levels of eye contact than persons with lesser degrees of social skill. During the experiment, two observers were stationed behind each of a pair of observation mirrors. One of each pair of observers recorded the time that a subject spoke, while the other member of the pair recorded the time that the subject looked at his or her partner. The observer pressed a key when the behavior of interest occurred, and the four keys operated the pens of an event recorder. Although this study demonstrates the use of a human servo technique, no experimenter accuracy or reliability scores are available.

A similar study to the one performed by Cherulnik et al. (1978) was undertaken by Argyle and Graham (1976). These researchers were interested in investigating the conditions under which greater or lesser amounts of gaze are directed at either a second person, objects placed between the two individuals, or the background. Two observers were seated behind a one-way window, with one observer monitoring each subject. Each of the observers measured the gaze of the subject seated diagonally across from him using a two-channel event recorder. One channel was used to record the duration of gaze at the

other person, and the second channel was used to record the duration of gaze at the object on the table. The human servo technique proved to be extremely useful in gathering data on subject behaviors. Finally, in a study by Lindman (1980), the effects of alcohol on mood and eye contact were investigated. Once again, subject behaviors (eye contact and averted gazes) were successfully recorded by observers situated behind an observation mirror via electronic timing devices.

Although all of these researchers took advantage of the human servo technique in measuring behavior durations, there is one distinct difference between their use of the technique and the use of the technique to measure in-car manual task durations. None of the studies cited above involved a warning stimulus or foreperiod. Observers were required to press some sort of hand-control at the instant they were exposed to the stimulus of interest, without any type of preparation. Because of the uncertainty involved in the use of the human servo technique in such a manner, it would seem logical to assume that simple reaction times were longer in these studies than they would be in an in-car task recording situation.

However different these studies are from the scenario established at the beginning of this literature review, they do provide some insight into what can be expected of the timing analyst. Based upon the results from the Johansson and Rumar study (1971), it appears that within-observer simple reaction time variability can be controlled through proper selection, training, and “calibration” of the observer. As for interrater reliability, the Goldberg and Wellens (1979) study leads one to predict very high correlations between results obtained from various observers.

General conclusions regarding the human servo technique. In conclusion, it appears that the human servo technique can be very useful for measuring in-car manual task durations. Within-observer simple reaction time variability appears to be both controllable

and minimizable, and the correlations between results obtained by different observers can be expected to be high. As far as between-observer simple reaction time variability is concerned, it would not be surprising if this is slightly elevated due to individual psychophysiological differences.

Errors of measurement can be expected to occur when using the human servo technique as described in the predefined scenario to measure in-car manual task durations. A reaction time delay can be expected at the beginning of a task duration measurement, however this delay appears to be somewhat controllable through the use of slow, carefully worded instructions. As far as the reaction time delay occurring at the end of a task duration measurement is concerned, the bulk of the research evidence seems to suggest that it will be somewhat larger than that at the beginning of a task. Fortunately, any errors in measurement that might occur as a result of simple reaction lags or movement time lags will probably be negligible when compared to the amount of time an in-car task demands.

Frame-By-Frame Technique

The frame-by-frame technique provides another way of measuring driver in-car manual task durations. As in the discussion of the human servo technique, it is first necessary to describe a common scenario within which this procedure is used. Once again the scenario includes a subject (driver) and a front-seat experimenter seated to the right. In this scenario, however, there is no back-seat timing analyst responsible for observing the driver as the tasks are performed. A specially mounted video camera and recording unit are used for that purpose. The recordings made from this system are later examined via frame-by-frame analysis (described earlier) to determine the durations of the in-car manual tasks.

Potential problems. Since the frame-by-frame technique for recording in-car manual task durations is so objective (in comparison to the human servo technique), many of the

problems encountered with the previously described technique can be avoided. Unfortunately, there are still opportunities for error with this system.

A modern video camera records activity at a speed of 30 frames per second. As a result, each frame of the recording represents one-thirtieth second of real time. Furthermore, the amount of time represented between successive frames in the recording is equal to one-thirtieth second. In almost all situations, a behavior will begin prior to the point represented by the first frame of videotape portraying that presence of that behavior. Furthermore, the behavior will almost always end prior to the point represented by the first frame of videotape portraying the absence of that behavior. If the behavior is completely symmetrical, i.e. it begins and ends equally early, there will be no error in the measurement of its duration. If, on the other hand, the behavior is not completely symmetrical, there is the possibility of a maximum of one-thirtieth second of error in the measurement of its duration.

The entire probability distribution of error due to sampling can be found by convolving the random error probability density functions at either end of the time interval of interest. First, a probability density function of the random timing error occurring at the beginning of the behavior interval is developed. Figure 3 illustrates this distribution. Next, a probability density function of the random timing error occurring at the end of the behavior interval is developed. Figure 4 illustrates this distribution. Finally, the convolution integral is used to derive a probability density function of the total timing error due to sampling. This general probability density function accounts for all errors resulting from the overlap of the two individual error probability density functions, and is shown in Figure 5. The mean value of this distribution is zero, corresponding to an absence of errors due to sampling. In addition, the variance is 0.00019 s^2 , and the standard deviation is 0.01361 s .

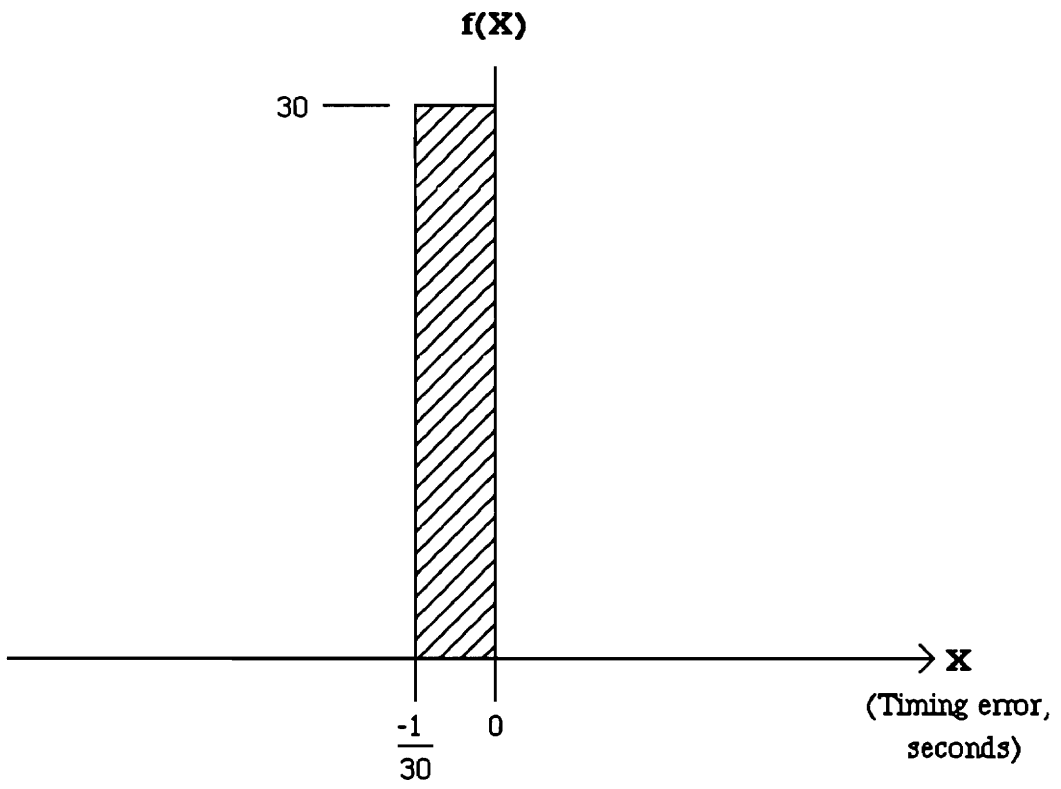


Figure 3: Probability density function of the random timing error (X) occurring at the beginning of the behavior interval.

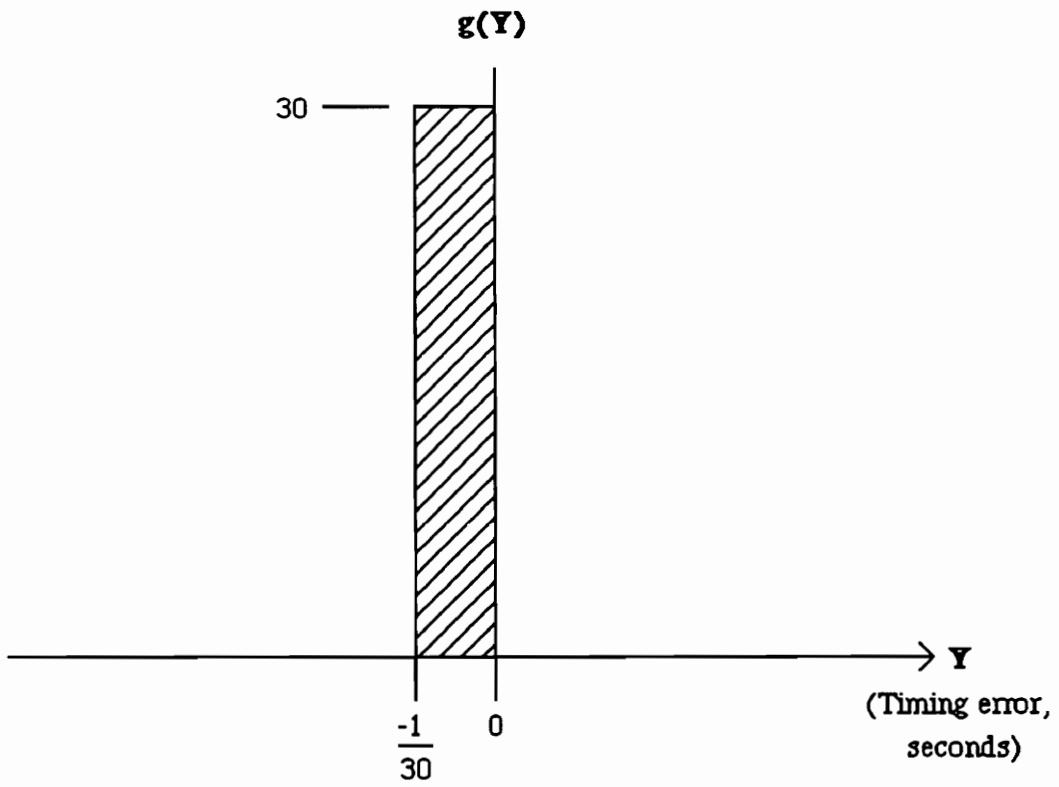


Figure 4: Probability density function of the random timing error (Y) occurring at the end of the behavior interval.

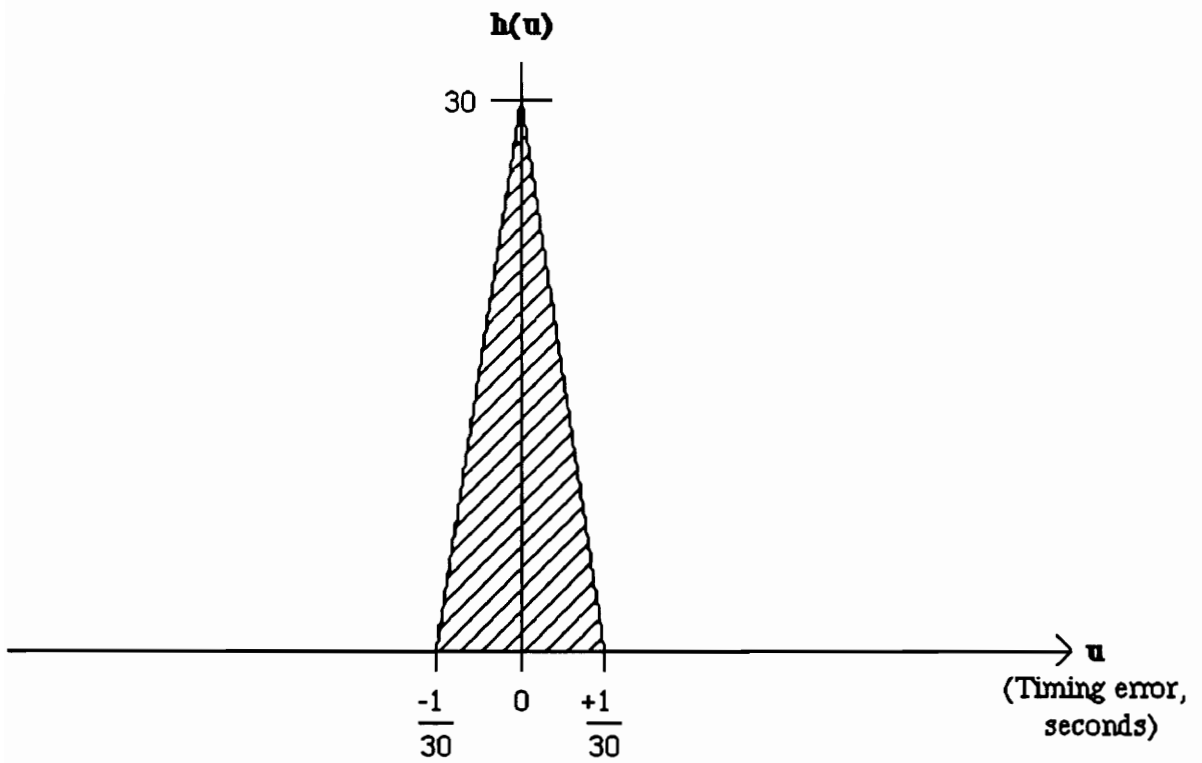


Figure 5: General probability density function of the random timing error (u) due to sampling.

Obviously, any errors that do occur as a result of sampling will be rather small. In fact, they will never be larger than one-thirtieth second, as demonstrated above. When one compares the length of these errors to the duration of a typical in-car manual task, they appear to be inconsequential.

Additional problems may exist when using the frame-by-frame technique to analyze driver in-car manual task durations. For one, there is always the possibility that a driver might perform a task in such a manner that the camera cannot capture what is actually happening. As an example, the driver's arm might block the view of his or her hand when the hand is taken off the steering wheel or when it is returned. Furthermore, the camera resolution may not be sharp enough or the background lighting may be too intense to discern the fine details of what a driver is doing, making subsequent analysis of an in-car task extremely difficult to perform. Finally, there is the possibility that an observer examining a videotape via frame-by-frame analysis may suffer from fatigue and may consequently report erroneous task durations (G. B. Stewart, personal communication, June 13, 1991).

General conclusions regarding the frame-by-frame technique. The frame-by-frame technique for measuring in-car manual task durations is more objective and possibly more accurate than the human servo technique. Unfortunately, errors still occur and new problems are introduced that were not at issue with the human servo technique. Generally speaking, the errors should be insignificant relative to the amount of time represented by a typical task.

Because the frame-by-frame analysis of video-recorded data is quite objective (although not *perfectly* objective), it is speculated that there is no problem with interrater reliability. Considering how high the reliability coefficients were in the Goldberg and Wellens (1979) study, one would tend to assume that they would be at least as high for the

frame-by-frame technique. Interestingly, most error variance associated with the video recording technique is outside of the control of the timing analyst. Hence, the relative accuracy of the method is apparent.

Alternate Task Duration Measurement Techniques

Two additional ways of implementing the human servo technique exist that may reduce or eliminate many of the problems described earlier. In the first technique, a videotape of the in-car manual tasks is observed in real time by a timing analyst, and the tasks are timed with a hand-held pushbutton control. Using this technique, the observer can always go back and re-time specific behaviors of interest if the original results appear unreasonably inaccurate. In the second technique, a videotape of the in-car tasks is observed by the timing analyst as described above, but in slow motion. By manually timing in-car tasks in this way, simple reaction time delays may be reducible relative to task durations.

One final technique for measuring the durations of in-car manual tasks deserves mention. If the steering wheel is outfitted with a touch-sensitive strip, the strip can serve as an on/off switch for an electronic timer. When the driver takes a hand off the strip, the timer can be activated. Similarly, the timer can be stopped when the driver places the hand back onto the strip. Using this technique, the timing analyst does not have to be concerned with errors introduced by simple reaction time lags or inaccuracy of a video-recording system. This technique could possibly be the most objective one of all for measuring in-car task durations, if everything were to work as planned.

Of course, every technique has its limitations, and even this last one is not exempt. If the driver takes a hand off the touch-sensitive strip for even a moment without intending to perform a secondary task, the timer will be falsely activated. In a similar vein, unintended

contact with the strip prior to completion of the secondary task may falsely deactivate the timer. There is also the possibility of activating the sensor strip with the wrong hand, and this can lead to considerable errors. Thus, even the most seemingly objective procedure may suffer from errors.

Summary

In summary, previous research suggests that the frame-by-frame technique for measuring in-car manual task durations is more accurate than the on-road version of the human servo technique (the scenario used in this literature review). While the frame-by-frame technique does suffer from framing (time-sampling) errors, simple reaction time research predicts that the on-road human servo technique suffers from larger errors as a result of observer simple reaction time delays. Unfortunately, even though the frame-by-frame technique is only subject to minimal errors, it requires the investment of more money, equipment, and time than the human servo technique.

A thorough analysis of several variations of the human servo technique could reveal whether any of these variations approach the preciseness of the frame-by-frame technique. At the same time, such an analysis could show whether one version of the human servo technique is any more accurate than another version. If it could be demonstrated that any variations of the human servo method approach the frame-by-frame technique in terms of preciseness, the future use of such variations could result in substantial savings in experimental resources.

Furthermore, it would be interesting to determine whether the lengths of in-car tasks have any effect on the accuracy of task duration measurements. Perhaps there is an effect here that is not predicted by the literature. Along similar lines, there may be a general subject effect on the accuracy of in-car task duration measurements. Such results would

also impact future transportation-related studies to a considerable degree. The following research was designed to address these issues.

METHOD

Subjects

Six subjects (three males and three females) volunteered to participate in this study. All subjects were graduate students in the human factors engineering program at Virginia Polytechnic Institute and State University, chosen because of their familiarity with reaction time theory and human factors methodology. They were typical of the the type of people who would be serving as timing analysts in an actual case. Subjects were screened using a test of visual acuity. The purpose for this screening was to ensure that they could adequately see the starting and stopping points for each in-car manual task performed by the driver.

Each subject was paid \$5.00 per hour for participation in this experiment. The study took approximately four hours for each individual to perform (a little over one hour on the road, and a bit over three hours in the Vehicle Analysis and Simulation Laboratory at Virginia Polytechnic Institute and State University). Thus, a total payment of approximately \$20.00 per subject was issued.

Apparatus

A Titmus II Vision Tester, manufactured by Titmus Optical, Inc., was used to screen each subject for acceptable near (14 inches) and far (20 feet) visual acuity. The vision test involved the use of a binocularly presented version of the Landolt Ring Test. For both the near and the far tests of acuity, subjects were required to demonstrate 20/40 vision or better.

The research vehicle used in this study was a 1989 Oldsmobile Cutlass Supreme with an automatic transmission and a vacuum fluorescent instrument cluster. This vehicle is shown in Figure 6. The vehicle contained two types of control systems for performing



Figure 6: Research vehicle used in this study -- 1989 Oldsmobile Cutlass Supreme.

radio-based tasks and HVAC-based tasks (heating, ventilation, and air-conditioning). One set of controls was mounted on the center of the steering wheel, called the steering *hub*, and the second set of controls was positioned on the right-hand portion of the instrument panel. These two control systems were fairly redundant.

Two Sony CCD video cameras mounted inside the research vehicle were used to record the driver's hand-off-wheel times as in-car tasks were performed. One camera, positioned above the driver, was focused on the steering wheel and the driver's hands. The second camera was positioned above and to the right of the driver, and it was focused on right-hand portions of the instrument panel. These cameras were connected to a compact Sony VHS video cassette recorder through a quad-splitter. The video cassette recorder was used to record the camera images onto a single videotape. A time code generator installed in the vehicle was used to insert a digital date and time stamp on each frame of the resulting video recording. Furthermore, both the driver and the front-seat experimenter wore lapel microphones so that all commands and comments could be recorded onto the videotape.

Two LEDs positioned over the driver's right shoulder appeared on the video recording. The first LED was lit when the experimenter pressed a console-mounted pushbutton at the onset of a command. The second LED was lit (and the first was extinguished) when the experimenter pressed another pushbutton at the end of the command. Finally, the second LED was extinguished when the experimenter pressed a third pushbutton upon the return of the driver's hand to the steering wheel. These LEDs served solely as markers for quickly locating a task when using the frame-by-frame technique.

There were several extraneous pieces of hardware in the research vehicle during the execution of this experiment. These included two additional CCD cameras, a second VHS

video cassette recorder, and a personal computer. These auxiliary hardware units were all mounted in the automobile for use in another study (although the additional cameras automatically recorded the forward road scene and the driver's eye movements onto the quad-split videotape while the experiment was being run). The vehicle was also equipped with a fire extinguisher, a first-aid kit, and a CB radio which could be used to contact public safety authorities in the event of an emergency.

Human servo timing of in-car manual tasks was performed with the aid of an electronic stopwatch. This LED stopwatch, manufactured by Cronus Precision Products, Inc., was accurate to one hundredth second. When pressed less than one-tenth millimeter, a pushbutton control on top of the stopwatch initiated the timing process. When pressed again, the same pushbutton control terminated the timing process. Elapsed time was displayed on the face of the device. A reset button on the front of the stopwatch, below the elapsed time display, was used to clear the current display and prepare the stopwatch for the next timing sequence. Figure 7 portrays the Cronus 4 LED electronic stopwatch.

Video analysis segments of the study were performed in the Vehicle Analysis and Simulation Laboratory using a JVC Super VHS stereo video cassette recorder and a 20-inch Sony Trinitron color monitor. This system was used to playback recordings of driver task performance at regular speed, one-sixth regular speed, and frame-by-frame speed to determine task durations. Figures 8 - 10 illustrate the video analysis system.

An IBM personal computer, model number 5150, was used to analyze the data from this experiment. This computer contains 256K of random access memory and is connected to a mainframe computer via a private telephone line. The SAS (Statistical Analysis Software) computer software program, version 6.06.01 for the mainframe, was used to perform statistical analyses of the resulting data.

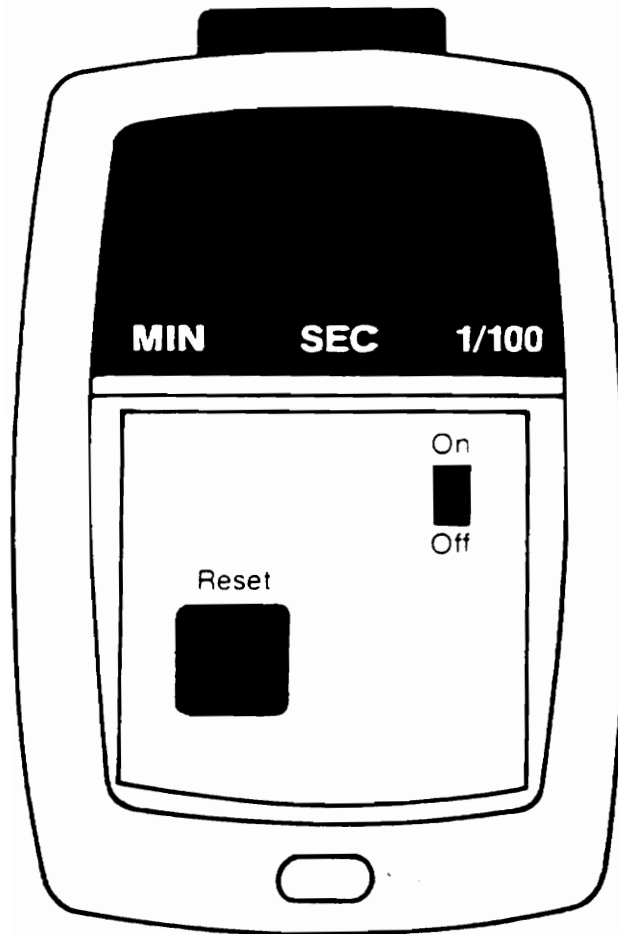


Figure 7: Cronus 4 LED electronic stopwatch.



Figure 8: Video analysis station.



Figure 9: Video quad image viewed while measuring in-car manual task durations with the video servo techniques and the frame-by-frame technique. Steering wheel/hub controls are pictured at the bottom left, and dash controls are pictured at the bottom right (top two images were not analyzed).



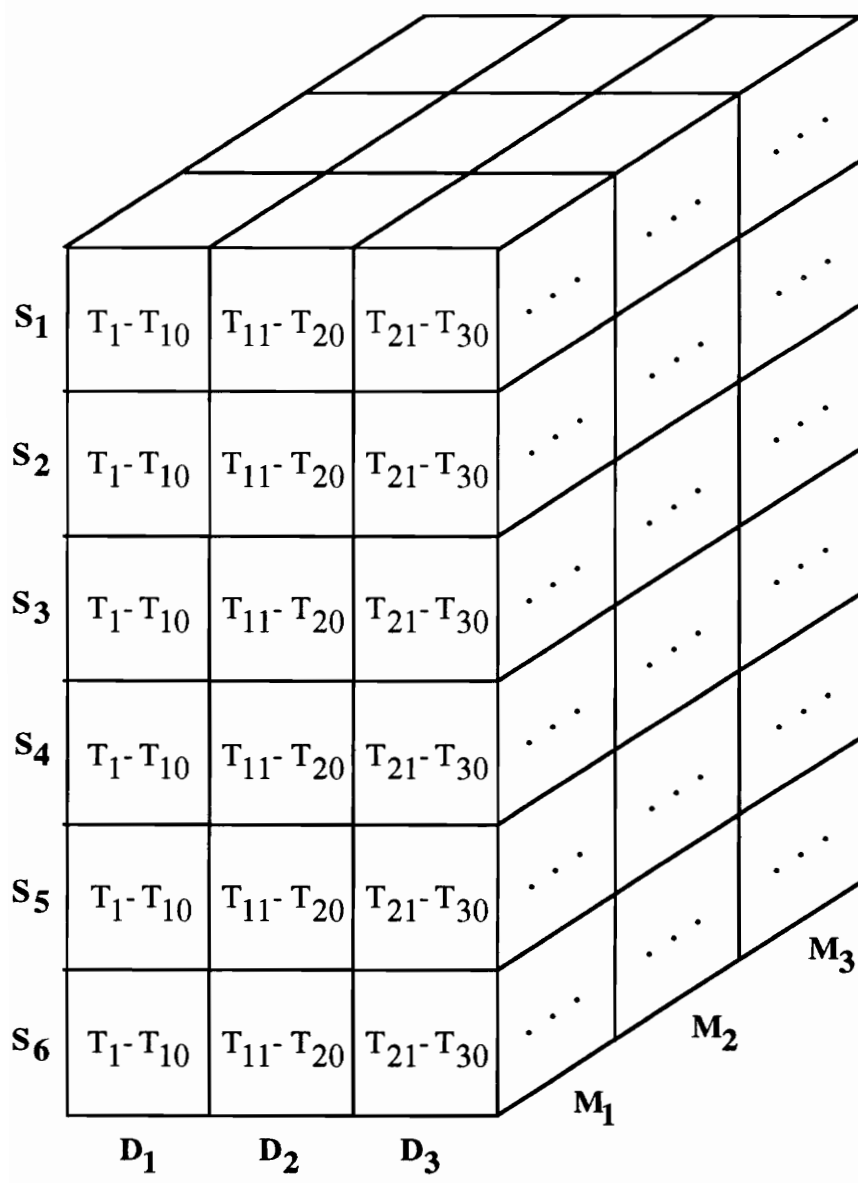
Figure 10: Close-up view of VCR remote control used during video analysis segments of the study.

Experimental Design

The experimental design used in this study was a mixed-factor, 6 x 3 x 3 complete factorial. The three factors were Subject, human servo Method, and in-car manual task Duration. Because Subject (six levels corresponding to six subjects) served as one of the independent variables, each cell of the experimental design contained ten “replications” of a particular type of in-car manual task. Thus, rather than Task being a within-*subject* variable, Subject was a within-*task* variable. Method was also a within-*task* variable, and it contained three levels: on-road servo, real-time video servo, and slow-motion (“slo-mo”) video servo. As for Duration, it was a between-*task* variable containing three levels: short tasks (expected to take less than 5 seconds to perform), medium tasks (expected to take between 5 and 10 seconds to perform), and long tasks (expected to take greater than 10 seconds to perform). The experimental cell design is illustrated in Figure 11.

This study contained three dependent variables. The first dependent variable was raw performance error, a measure of response *bias*. The operational definition of raw performance error is the difference between the duration measurement made by a subject for a given task and the “true” duration of that task based upon frame-by-frame analysis. There were ten of these error scores in each experimental cell, summing to 30 in-car tasks that a subject was required to time using each of the three human servo methods. Since each subject had to use all three human servo methods, and six subjects participated in the study, the total number of data points was 540.

The second dependent variable was absolute performance error. This measure of response *accuracy* was obtained by taking the absolute value of each of the 540 raw performance error data points. Finally, the third dependent variable was squared performance error. This is a measure of response *consistency*. The data corresponding to this variable were derived by squaring each of the 540 raw performance error data points.



KEY: S = Subject, D = Duration, M = Method, T = Task

Figure 11: Experimental cell design.

The three dependent variables will be discussed in more detail in the Results section of this paper.

Each of the six subjects was randomly assigned to a Subject level prior to the collection of any data. As a result of this random assignment, Subjects 2, 3, and 5 were males and Subjects 1, 4, and 6 were females.

The presentation order of both within-task factors was designed to avoid any differential transfer effects. A completely balanced latin square design was used to counterbalance presentation order of the three human servo Method levels across the six Subject levels. This design was composed of a standard latin square design for the first three subjects (1, 2, and 3) and a non-standard latin square design for the second three subjects (4, 5, and 6). Table 1 shows how counterbalancing of human servo Method presentation order was accomplished. Critical path charts used to schedule subjects based upon this counterbalancing scheme are portrayed in Figure 12. Since there was at one point a possibility of only having time to use three subjects, the first three subjects were completely run before the second three subjects were run.

Differential transfer effects of the in-car task Duration factor were controlled through the randomization of task presentation order for each on-road run. In other words, tasks were randomly chosen from the three different levels of the Duration factor and randomly assigned to a presentation position for every run. This design eliminated the need for any sort of complicated counterbalancing scheme.

Procedure

Experimental tasks. Each subject in this study measured in-car manual task durations using three different techniques. The first technique (on-road servo) involved measuring in-car task durations while serving as an observer in the back seat of the

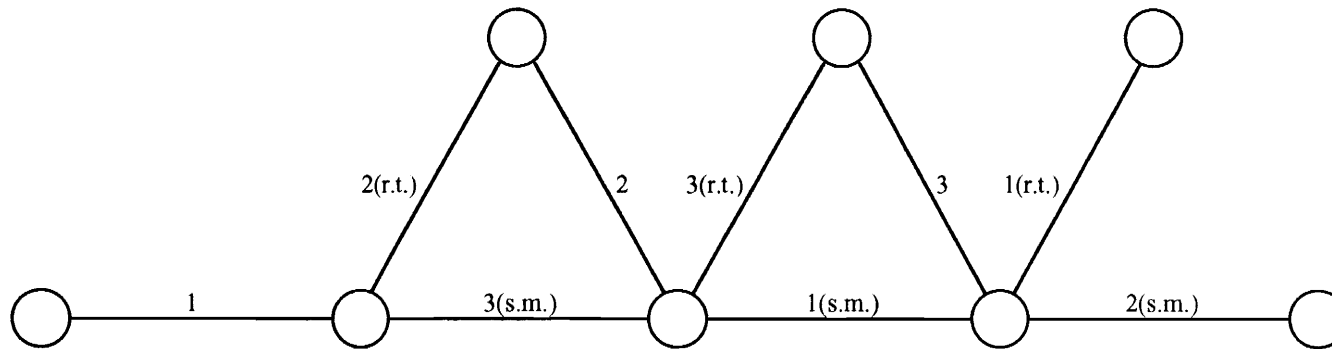
Table 1

Counterbalancing of human servo Method presentation order.

<i>BLOCK 1</i> <i>Standard</i> <i>Latin Square</i>	ON-ROAD SERVO (REAL-TIME)	VIDEO SERVO (SLOW-MO)	VIDEO SERVO (REAL-TIME)
SUBJECT 1	1	2	3
SUBJECT 2	2	3	1
SUBJECT 3	3	1	2

<i>BLOCK 2</i> <i>Non-standard</i> <i>Latin Square</i>	ON-ROAD SERVO (REAL-TIME)	VIDEO SERVO (SLOW-MO)	VIDEO SERVO (REAL-TIME)
SUBJECT 4	1	3	2
SUBJECT 5	2	1	3
SUBJECT 6	3	2	1

NOTE: In the above counterbalanced blocks, the numbers 1 - 3 represent the within-subject presentation order of the Human Servo Method levels.



KEY:

1 = Subject 1, on-road servo

2 = Subject 2, on-road servo

3 = Subject 3, on-road servo

1(r.t.) = Subject 1, video servo (real-time), 3's on-road run

2(r.t.) = Subject 2, video servo (real-time), 1's on-road run

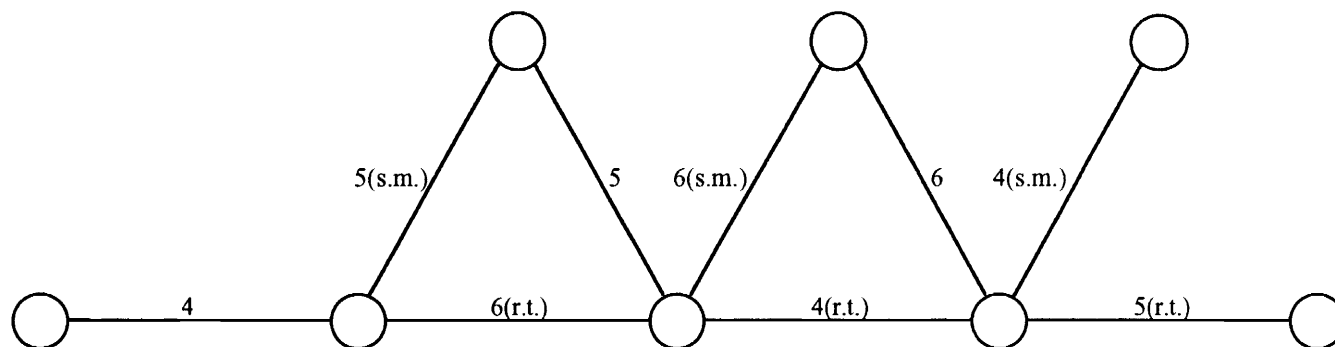
3(r.t.) = Subject 3, video servo (real-time), 2's on-road run

1(s.m.) = Subject 1, video servo (slo-mo), 2's on-road run

2(s.m.) = Subject 2, video servo (slo-mo), 3's on-road run

3(s.m.) = Subject 3, video servo (slo-mo), 1's on-road run

Figure 12: Critical path charts for scheduling subjects.



KEY:

4 = Subject 4, on-road servo

5 = Subject 5, on-road servo

6 = Subject 6, on-road servo

4(r.t.) = Subject 4, video servo (real-time), 5's on-road run

5(r.t.) = Subject 5, video servo (real-time), 6's on-road run

6(r.t.) = Subject 6, video servo (real-time), 4's on-road run

4(s.m.) = Subject 4, video servo (slo-mo), 6's on-road run

5(s.m.) = Subject 5, video servo (slo-mo), 4's on-road run

6(s.m.) = Subject 6, video servo (slo-mo), 5's on-road run

Figure 12 (continued)

research vehicle. The second technique (real-time video servo) involved timing in-car tasks while watching a real-time videotape of driver performance in the Vehicle Analysis and Simulation Laboratory. The third technique (slow-motion video servo) required subjects to time in-car manual tasks while watching a videotape of driver performance at one-sixth speed in the laboratory. As shown in Table 1, these three conditions were counterbalanced across subjects to avoid any differential transfer effects.

For the on-road servo technique, the subject sat in the back seat of the research vehicle while the driver and an experimenter sat in the front seat. The front-seat experimenter was stationed in the passenger seat to the driver's right, and the subject was positioned directly behind the experimenter. During an experimental run, the experimenter was responsible for (1) providing verbal commands for the driver to follow; (2) activating console-mounted buttons (controlling the LED lights) at the beginning of a verbal command, the end of the verbal command, and the completion of the task; (3) setting controls to appropriate positions between commands; (4) manning an additional safety brake; and (5) ensuring that the vehicle was safely operated. The subject was required to (1) measure the hand-off-wheel time for each of the in-car manual tasks using an electronic stopwatch and (2) record these hand-off-wheel times on a log sheet (See Appendix A).

The experimenter gave all commands slowly and clearly. The driver began all tasks *upon completion of the instructions for that task* and kept both hands on the steering wheel at all times that a task was not being performed. As mentioned earlier, each of the 30 tasks performed by the driver was classified into one of three categories, depending on the expected duration. There were ten tasks presented under each of the three duration categories during an experimental run, and these tasks were randomly distributed. Each task was separated in time by a one-minute rest period. The 30 in-car manual tasks are presented in Appendix B.

The subject was instructed to initiate timing of a task at the instant he or she saw the driver's hand "lose contact" with the steering wheel. As soon as the subject saw the driver's hand lose contact with the wheel, that subject pressed the button on the top of the handheld stopwatch. The subject was told to "react and move as quickly as possible" whenever activating the pushbutton control. When the driver completed the task and the hand "made return contact" with the steering wheel, the subject once again pressed the stopwatch timing button. The elapsed time appearing on the face of the stopwatch was then recorded by the subject on a log sheet, and the reset button was pressed to prepare the stopwatch for the next trial. If the subject made an error in timing the task (as one subject did), the error was recorded on a separate error log sheet (Appendix A).

All on-road trials were performed along a stretch of U.S. Highway 460 between Blacksburg and Pembroke, Virginia. This is an improved, divided route with good visibility and low-to-moderate traffic density. Experimental runs took place only in weather that did not adversely affect visibility or road adhesion. Trials were performed while running both to Pembroke from Blacksburg and back.

Five practice tasks were performed at the very beginning of each on-road run, so that the subject could become acquainted with the timing technique and could ask any questions that may have arisen. These five tasks were chosen to represent the range of tasks that the subject would be required to time. They are presented in Appendix B along with the 30 in-car manual tasks.

For the real-time video servo technique, subjects were brought into the Vehicle Analysis and Simulation Laboratory and seated in front of the video analysis system. A videotape of driver task performance from an unfamiliar research vehicle run was shown to the subject at normal playback speed, and that subject timed and logged all 35 tasks that the driver performed (once again, the first five tasks were practice trials). The basic timing

procedure was the same as that described for the on-road trials. In addition, the subject could rewind the videotape whenever he or she made a mistake.

For the slow-motion video servo technique, subjects again had to time tasks while stationed in the Vehicle Analysis and Simulation Laboratory. As with the real-time video servo technique, a videotape of driver task performance from an unfamiliar research vehicle run was shown to the subject. However, the subject was required to switch the video cassette player to one-sixth regular speed when the verbal instructions preceding a task were coming to an end. While playing the tape at one-sixth regular speed, subjects timed and logged the five practice tasks and 30 experimental tasks that the driver performed. The subject could rewind the videotape whenever he or she made a mistake and play the tape at normal speed to move quickly from the end of one task to the end of the instructions for the next.

A subject was not exposed to the same research vehicle run more than once during the course of his or her participation in this study. This means that the two videotapes played for the subject were each recorded from different on-road trials than the one in which the subject had participated or was to participate. Additionally, even though the 30 tasks were the same under each of the three timing conditions, their presentation order differed from condition to condition. Only the five practice trials were identical in presentation order from condition to condition.

There was only one driver for all six of the on-road experimental sessions. This driver was responsible for operating the research vehicle and performing the in-car tasks. The use of a single driver eliminated any confounding effects that could have arisen as a result of having several individuals with different experience levels or driving styles operate the automobile and perform the required tasks.

An additional experimenter analyzed the six videotapes from the on-road trials using the frame-by-frame technique. For each of the 35 tasks on a videotape (including the five practice tasks), the experimenter recorded (1) the time code for the first frame in which the driver's hand was no longer in contact with the steering wheel, and (2) the time code for the first frame in which the driver's hand made return contact with the steering wheel. The first time reading was then subtracted from the second time reading to obtain the "true" measure of task duration. These true durations were recorded on a log sheet (see Appendix A) for comparison with the task durations obtained using the three different human servo methods.

Experimental protocol. Each subject participated in three separate sessions over three days. The three days were not necessarily consecutive. On the first day, the subject reported to the Vehicle Analysis and Simulation Laboratory, where he or she was immediately screened for both near and far visual acuity. If the subject passed the vision test, he or she was asked to read the general instructions for the experiment and the informed consent form. These forms are displayed in Appendices C and D, respectively. The subject then signed the informed consent form if he or she agreed to the conditions of the study. Once the subject signed the informed consent form, the experimenter verbally reviewed the instructions for the first experimental session and answered any questions the subject may have. The first session required the subject to time driver performance of in-car manual tasks using either the on-road servo technique, the real-time video servo technique, or the slow-motion video servo technique, depending upon the subject's position in the counterbalancing scheme.

The remaining sessions followed the same basic protocol. After the entire experiment had been completed, the subject was debriefed and was given the opportunity to ask any

questions that may have arisen during the course of the study. The subject was then paid for the time spent performing the experiment.

Research Hypotheses

For both the raw-error-score ANOVA and the absolute-error-score ANOVA, it was hypothesized that the human servo Method effect would be significant (the results of the squared-error-score ANOVA were unpredictable, so no research hypotheses were formulated). Specifically, use of the slow-motion video servo technique was expected to result in raw error scores approximating zero, and use of the two real-time servo techniques was expected to result in significantly greater raw error scores (more positively biased).

There are two primary reasons why the raw error scores for the slow-motion video servo technique were expected to be so small. First, any simple reaction delays which might be produced by subjects in the timing of slow-motion tasks would be reduced to one-sixth their values when the duration measurements were converted to real time. Simple reaction foreperiod research suggests a second reason why the raw error scores for the slow-motion video servo technique were expected to be close to zero. At the end of a slow-motion video servo task, the timing analyst's simple reaction foreperiod (preceding the return contact of the driver's hand) will generally be considerably longer than the simple reaction foreperiod occurring at the end of a real-time task (on-road or video). This is due to the longer time period between the driver's hand leaving the task controls and making return contact with the steering wheel. According to studies performed by Drazin (1961), Mo and George (1977), Nickerson (1965), and Polzella et al. (1989), simple reaction time will decrease with foreperiod length when variable foreperiods are involved. Since the ending foreperiods for the tasks in this study *were* variable in length, simple reaction times at the ends of the slow-motion tasks were expected to be smaller than those at the ends of

the real-time tasks. Furthermore, simple reaction times at the beginnings of the slow-motion tasks were also expected to be small, since the subjects were permitted to set the video cassette player to slow speed toward the end of the initial foreperiod. In essence, once the subjects set the video cassette player to one-sixth real speed, they knew that the reaction stimulus was soon to follow and were prepared for it. Assuming that the beginning and ending simple reaction delays were to be short and relatively equal in duration, there were not expected to be any considerable bias or accuracy problems associated with using the slow-motion video servo technique.

As stated above, use of the on-road servo technique and the real-time video servo technique was expected to result in significantly greater raw error scores than those found through use of the slow-motion video servo technique. This result was anticipated for two primary reasons: (1) there was to be no reduction of simple reaction delays to one-sixth their values once all of the data had been collected, and (2) the variable foreperiod research cited in the previous paragraph suggests that simple reaction times should have been longer at the ends of tasks timed using these techniques than at the ends of tasks timed using the slow-motion technique. For the real-time techniques, the ending simple reaction times were expected to be longer than the beginning simple reaction times, since the ending foreperiods were considerably shorter than the beginning foreperiods. As a result, true task durations were expected to be overestimated when using these two techniques.

Two studies on foreperiod length led to the hypothesis that there would be a significant task Duration effect for both the raw error data and the absolute error data. Specifically, it was hypothesized that the short tasks would lead to significantly greater performance error scores (more positively biased) than either the medium or the long length tasks. It is *possible* to consider the warning interval for the second pushbutton actuation in a set to start at the end of the first pushbutton actuation in the set. Hence, the observer of a

short task is provided with a fairly short warning interval (in comparison with medium and long tasks) before the driver's hand makes return contact with the steering wheel. The previously mentioned research of Nickerson (1965) and Possamai et al. (1973) suggests that, as a result of such a relatively short warning interval, a subject might not have as much time to prepare to respond as he or she would have when observing a medium- or long-duration task. Therefore, simple reaction time could suffer. Although this problem might be alleviated somewhat by the ability of the subject to follow the driver's hand movement and predict the exact moment of return contact with the steering wheel (Simon and Slaviero, 1965), the warning interval is still relatively short enough to support the hypothesis that subject performance would differ. Again, no hypotheses were formulated for the squared error data.

A significant Subject effect was hypothesized for all three ANOVAs. Discussions by Boff and Lincoln (1988), Sanders and McCormick (1987), and many others (Huchingson, 1981; Johnson, 1923; Teichner, 1954; and Wickens, 1984) point to an assortment of factors which can affect a timing analyst's simple reaction times. Several of these factors are related to human performance, such as motivation and fatigue. Because the dependent variables in this study reflected errors attributable to simple reaction time delays, it was expected that individual subject performance differences would reveal themselves in the form of a significant Subject main effect.

It was also hypothesized that there would be a significant effect of Task for the raw-error-score ANOVA and the absolute-error-score ANOVA. This follows from the hypothesis of a significant Duration effect for the same two ANOVAs. Since Task is nested within the Duration factor, a significant Duration effect would mean that tasks of different durations would also be significantly different.

There was no concrete evidence to suggest that there would be any significant Subject x Method or Subject x Duration interactions in this study. It was hypothesized that, even though subjects would most likely differ from one another in terms of bias, accuracy, and consistency, each would exhibit the same general performance relationships across the three human servo techniques and across the three different task lengths. No hypotheses were formulated for the Method x Duration interaction or the Subject x Method x Duration interaction.

Finally, there was expected to be no significant effect of Gender on any of the dependent variables. This hypothesis was based on the fact that there is no known research suggesting that males are different from females in terms of simple reaction time. Additionally, the lack of any obvious strength or skill components required to use the human servo techniques would lead one to predict that there would not be a significant Gender effect.

RESULTS

When the 540 recorded times were examined, it was discovered that some of the medium-length tasks had taken the driver more than 10 seconds to perform. As a result, the definitions for medium- and long-length tasks had to be redefined so that all 54 cells in the experimental design would still contain an equal number of values. By setting the break-point between medium- and long-duration tasks at 12.15 seconds (rather than 10 seconds), only four medium tasks ended up being too long, and only one long task ended up being too short. It was decided that the four “long” medium tasks would still be considered to be of medium duration, and the one “medium” long task would still be considered to be of long duration. The final breakdown of Duration levels was as follows: short tasks took less than 5 seconds to perform, medium tasks took from 5 to 12.15 seconds to perform, and long tasks took more than 12.15 seconds to perform.

One data point was missing out of the final 540, since a subject using the on-road servo technique failed to stop the handheld timer at the end of a task. To avoid having to perform an unequal-cells analysis on account of only one missing data point, a substitute value was used. This substitute value was derived by averaging the recorded times for that same task across the remainder of the subjects. Although this substitution technique had the potential for reducing the variance between subjects and/or increasing the variance between tasks, there were enough data points to guard against any considerable effect.

The first step in analyzing the data was to convert all the recorded times into raw performance error scores. As mentioned earlier, these raw error scores were obtained by subtracting each task’s “true” duration (based upon frame-by-frame analysis) from the duration that was recorded by a subject. Thus, 90 raw error scores were derived for each subject (there were 30 tasks that were timed using each of the three human servo methods), and a total of 540 error scores were obtained across all six subjects. Because of the way

the error scores were calculated, a positive error score indicated an overestimation of true hand-off-wheel time and a negative error score indicated an underestimation of true hand-off-wheel time.

Three Analyses of Variance (ANOVAs) were conducted on the data. The first ANOVA compared the ten raw performance error scores in each experimental cell to determine if there were any biases in subject performance as a function of the independent variable(s) under consideration. Positive biases are indicative of a tendency to overestimate hand-off-wheel times, while negative biases are indicative of a tendency to underestimate hand-off-wheel times. In short, this analysis was used to establish the effect(s) of the variables of interest on the general direction of timing errors.

The second ANOVA compared the absolute values of the ten performance error scores in each cell. This measure made it possible to determine the accuracy of subject performance given all of the independent variable combinations of interest. Clearly, error scores which are close (or equal) to zero indicate accuracy of performance, whereas error scores which are greater or less than zero signify less-than-accurate performance. By concentrating on mean absolute error, this ANOVA took into account both the central tendency and the variance of the ten performance error scores in a cell.

Finally, the third ANOVA compared the squared values of the ten performance error scores in each experimental cell. This analysis was used to ascertain whether or not there was consistency in subject performance given different independent variables and combinations of these independent variables. The squaring process served to suppress the values of error scores which were close to zero and inflate scores which were farther away from zero. As a result, a relatively large squared error mean in a cell corresponded to relatively large timing variability, whereas a relatively small squared error mean signalled relatively inconsequential variability in timing of in-car tasks.

In ANOVAs which include random variables other than the dependent variable, certain main effects and interaction effects may, at first glance, seem unanalyzable due to missing denominators for their F ratios. In such cases, Quasi- F ratios can be constructed which combine several component mean squares to derive appropriate denominators. Since one of the independent variables analyzed in this study was random (Subject), Quasi- F ratios were used to test for main effects of Duration and Method, and for the interaction effect of Duration x Method. In addition, the Satterthwaite correction was used to adjust the tabled degrees of freedom for each Quasi- F ratio. Table 2 shows the component terms that were used to compose these ratios, as well as the equation used to calculate the tabled degrees of freedom (Satterthwaite correction) associated with each Quasi- F ratio. The ANOVA summary tables presented designate these Quasi- F ratios with an F' .

Post-hoc analyses of significant main effects and interaction effects were performed using the Newman-Keuls pairwise comparison technique (for the main effects) and the Tukey Honestly Significant Difference test (for the interactions). These “unplanned data-snooping” procedures were used to determine exactly how the independent variable levels affected the dependent variables. Additionally, post-hoc analyses were performed to determine if there was any Gender effect on subject performance error. These Gender analyses were performed using Student t -tests.

Analysis of Raw Error Scores

As shown in Table 3, the ANOVA performed on the raw error scores revealed a significant main effect of Method $F'(2,10) = 13.254, p = 0.002$. This effect indicates that timing analysts demonstrate differential biases depending upon which human servo technique they are using to time in-car manual tasks. Post-hoc analysis using the Newman-Keuls technique ($\alpha = 0.05$) revealed that the slow-motion video servo technique

Table 2

Construction of Quasi- F ratios and calculation of their tabled degrees of freedom.

<u>Quasi-F Ratio</u>	<u>Numerator Component</u>	<u>Denominator Component</u>
F'_{duration}	MS_{duration}	$(MS_{\text{task}}) + (MS_{\text{subject} \times \text{duration}}) - (MS_{\text{subject} \times \text{task}})$
F'_{method}	MS_{method}	$(MS_{\text{subject} \times \text{method}}) + (MS_{\text{method} \times \text{task}}) - (MS_{\text{subject} \times \text{method} \times \text{task}})$
$F'_{\text{duration} \times \text{method}}$	$MS_{\text{duration} \times \text{method}}$	$(MS_{\text{method} \times \text{task}}) + (MS_{\text{subject} \times \text{duration} \times \text{method}}) - (MS_{\text{subject} \times \text{method} \times \text{task}})$

Calculation of tabled degrees of freedom (Satterthwaite correction):

$$F' = (a) / (b + c - d)$$

$$df_{\text{numerator}} = df_a$$

$$df_{\text{denominator}} = (MS_b + MS_c - MS_d)^2 / [(MS_b^2 / df_b) + (MS_c^2 / df_c) + (MS_d^2 / df_d)]$$

Table 3

ANOVA summary table for raw error scores.

Source	df	SS	MS	<i>F</i>	<i>p</i>
<u>Between</u>					
Duration (D)	2	0.1118	0.0559	*	
Task/Duration (T/D)	27	1.6747	0.0620	2.48	0.0003
<u>Within</u>					
Subject (S)	5	0.2019	0.0404	1.61	0.1606
S x D	10	0.4370	0.0437	1.75	0.0767
S x T/D	135	3.3787	0.0250		
Method (M)	2	1.5118	0.7559	**	
M x D	4	0.0300	0.0075	***	
M x T/D	54	1.0951	0.0203	1.12	0.2791
S x M	10	0.5487	0.0549	3.03	0.0012
S x M x D	20	0.4962	0.0248	1.37	0.1372
S x M x T/D	270	4.8935	0.0181		
Total	539	14.3794			

* $F'(2, 19) = 0.693, p > 0.05$ ** $F'(2,10) = 13.254, p < 0.01$ *** $F'(4, 18) = 0.278, p > 0.05$

(mean = 0.078 s) resulted in significantly higher-valued errors than both the on-road servo and the real-time video servo techniques (means = -0.00006 s and -0.051 s, respectively). The mean timing error for each of the three human servo techniques, as well as the results of the post-hoc test, is depicted graphically in Figure 13 .

The raw-error-score ANOVA also uncovered a significant interaction of Subject x Method, $F(10, 270) = 3.03, p = 0.0012$. This interaction is illustrated in Figure 14 . Post-hoc analysis of the Subject x Method interaction was restricted to unconfounded pairwise comparisons to limit the inflation of alpha error as much as possible. To perform this analysis, the Cicchetti Table for Unconfounded Comparisons (Cicchetti, 1972) was used to adjust the Tukey Honestly Significant Difference test critical difference equation ($\alpha = 0.05$). The significant differences discovered by applying the Tukey Honestly Significant Difference test are presented in Table 4.

In addition to the aforementioned effects, the ANOVA performed on the raw error scores showed a significant effect of Task, $F(27, 135) = 2.48, p = 0.0003$. Mean timing error as a function of Task is presented in Figure 15. Since there were so many levels comprising the Task factor, post-hoc pairwise comparisons were not performed. However, it was decided that chi-square analyses should be performed to determine if there were any significant biases as a function of *type* of task that was timed (i.e., hub tasks, dash tasks, HVAC tasks, and radio tasks). The results of these chi-square analyses are presented in Table 5. These results show that (1) the hub-based task times were significantly biased in the negative direction, $\chi^2(1, N = 15) = 8.33, p < 0.01$, and (2) the dash-based task times were significantly biased in the positive direction, $\chi^2(1, N = 15) = 5.56, p < 0.02$. The chi-square analyses did not reveal any significant biases associated with the HVAC or radio tasks. Additionally, application of the binomial test revealed significant negative biases for hub-based, radio task times ($p = 0.016$) and significant

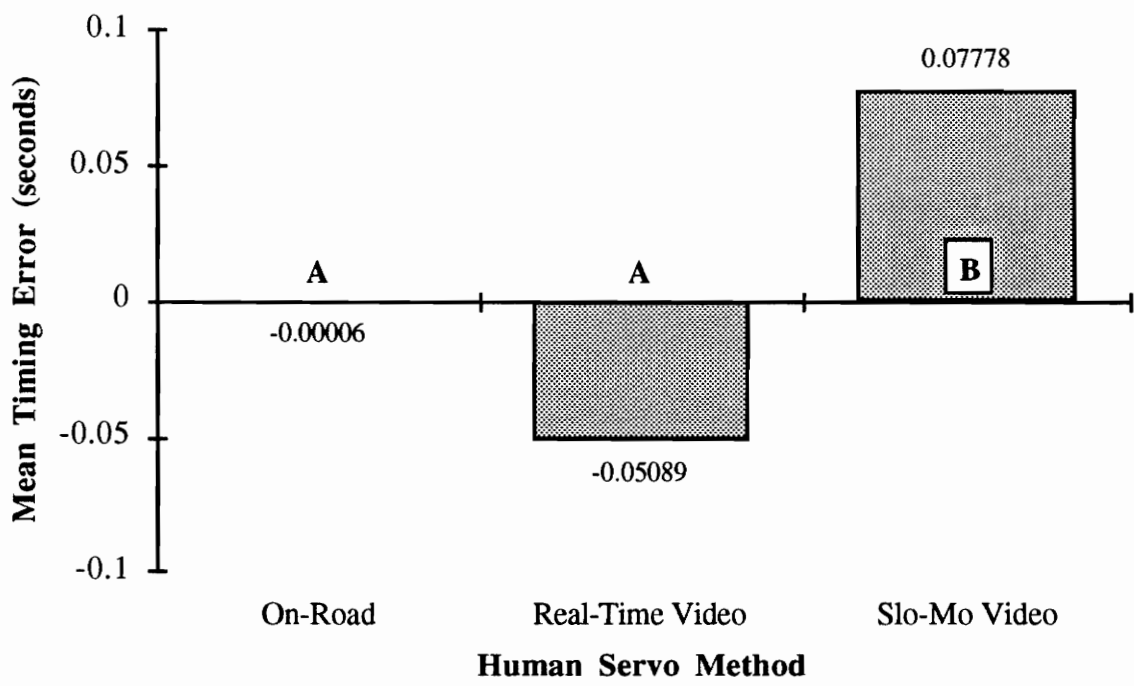


Figure 13: Mean timing error as a function of human servo Method. Means designated by common letters do not differ significantly ($\alpha = 0.05$).

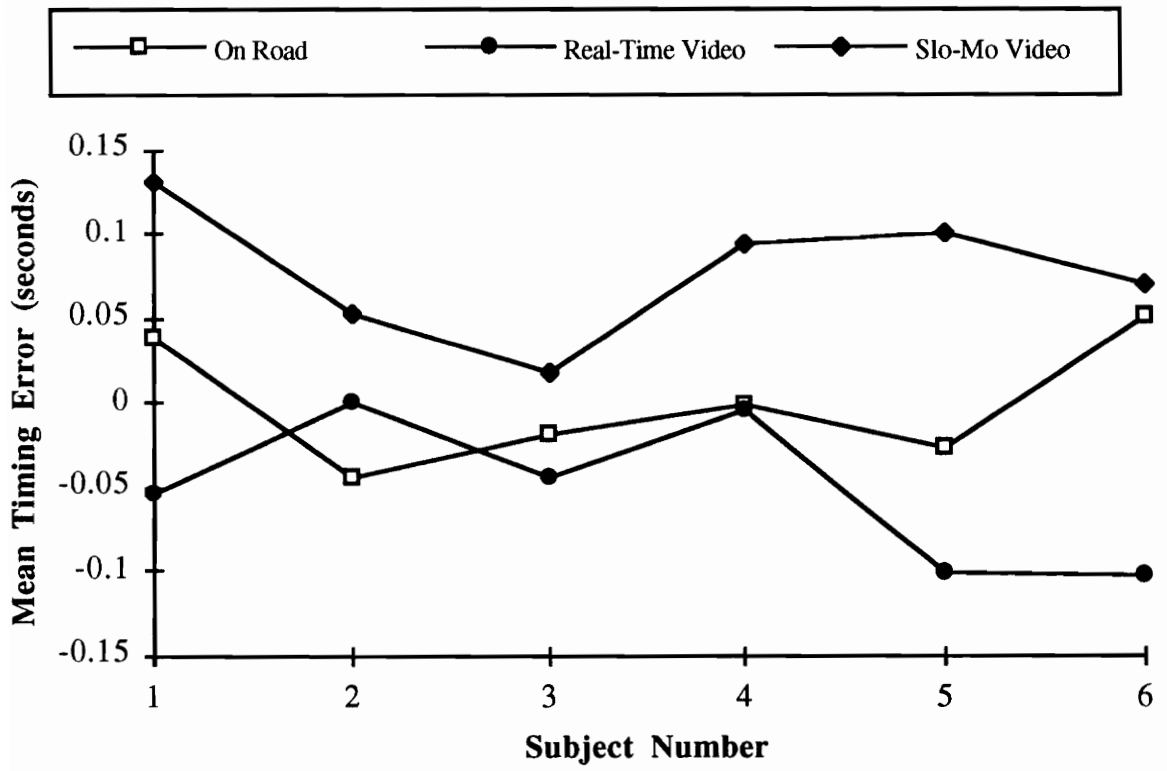


Figure 14: Subject x Method interaction (raw error data).

Table 4

Significant results of unconfounded pairwise comparisons for the raw-error-scores Subject x Method interaction, using the Tukey Honestly Significant Difference test.

Critical Difference _T = 0.11356 ($\alpha = 0.05$)			
Comparison			Difference Between Means
Subject 1: Slow-Motion Video (mean = 0.13133)	vs.	Real-time Video (mean = -0.05400)	0.18533
Subject 5: Slow-Motion Video (mean = 0.10133)	vs.	Real-Time Video (mean = -0.10133)	0.20267
Subject 5: Slow-Motion Video (mean = 0.10133)	vs.	On-Road (mean = -0.02667)	0.12800
Subject 6: Slow-Motion Video (mean = 0.06967)	vs.	Real-Time Video (mean = -0.10167)	0.17133
Subject 6: Real-Time Video (mean = -0.10167)	vs.	On-Road (mean = 0.05067)	0.15233
Slow-Motion Video: Subject 1 (mean = 0.13133)	vs.	Subject 3 (mean = 0.01767)	0.11367

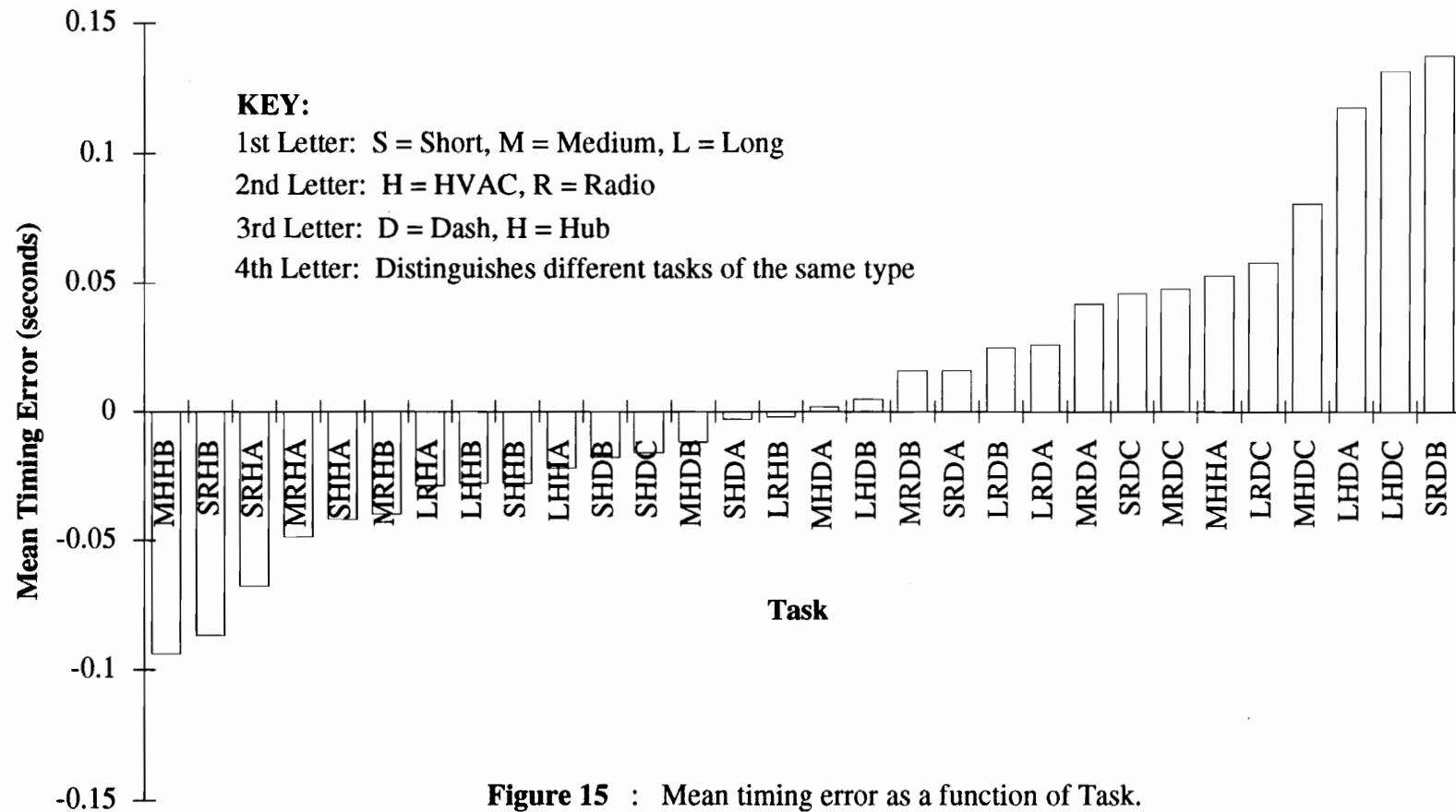


Figure 15 : Mean timing error as a function of Task.

Table 5

Results of chi-square analyses of Task means.

<u>HUB TASKS</u>			
	Observed Expected		
Number of Tasks with "+" Means	<table><tr><td>1</td><td>6</td></tr></table>	1	6
1	6		
Number of Tasks with "-" Means	<table><tr><td>11</td><td>6</td></tr></table>	11	6
11	6		

$\chi^2 = 8.33$
 $0.01 > p > 0.001$

<u>DASH TASKS</u>			
	Observed Expected		
Number of Tasks with "+" Means	<table><tr><td>14</td><td>9</td></tr></table>	14	9
14	9		
Number of Tasks with "-" Means	<table><tr><td>4</td><td>9</td></tr></table>	4	9
4	9		

$\chi^2 = 5.56$
 $0.02 > p > 0.01$

Table 5 (continued)

<u>HVAC TASKS</u>			
	Observed Expected		
Number of Tasks with "+" Means	<table><tr><td>6</td><td>7.5</td></tr></table>	6	7.5
6	7.5		
Number of Tasks with "-" Means	<table><tr><td>9</td><td>7.5</td></tr></table>	9	7.5
9	7.5		

$\chi^2 = 0.6$
 $0.5 > p > 0.3$

<u>RADIO TASKS</u>			
	Observed Expected		
Number of Tasks with "+" Means	<table><tr><td>9</td><td>7.5</td></tr></table>	9	7.5
9	7.5		
Number of Tasks with "-" Means	<table><tr><td>6</td><td>7.5</td></tr></table>	6	7.5
6	7.5		

$\chi^2 = 0.6$
 $0.5 > p > 0.3$

positive biases for dash-based, radio task times ($p = 0.002$). The results of the binomial tests are presented in Table 6.

Last, an a posteriori test of Gender, $t(538) = 2.28$, $p = 0.0232$, demonstrated a significant effect. The results of this Student t -test are presented in Table 7-a. Specifically, male subjects (mean = -0.007 s) made considerably lower-valued errors than female subjects (mean = 0.0232 s). Another way of looking at this is to say that males tended to slightly underestimate hand-off-wheel times, and females tended to overestimate hand-off-wheel times.

Analysis of Absolute Error Scores

As shown in Table 8, the ANOVA performed on the absolute error scores revealed a significant effect of Subject, $F(5,135) = 2.91$, $p = 0.016$. This result suggests that some individuals will tend to be more accurate than others when it comes to timing in-car manual tasks. Post-hoc analysis using the Newman-Keuls technique ($\alpha = 0.05$) revealed that Subject 1 (mean = 0.141 s) was significantly less accurate than Subjects 3 and 4 (means = 0.090 s and 0.081 s, respectively) when it came to measuring hand-off-wheel times. The Subject main effect is illustrated in Figure 16. No additional effects included in this ANOVA were significant. Furthermore, a post-hoc test of Gender failed to demonstrate any significant effect (Table 7-b).

Analysis of Squared Error Scores

Finally, the ANOVA performed on the squared error scores failed to demonstrate any significant main effects or interaction effects. In fact, the lowest p value in the entire analysis was for the Subject main effect, and this value was 0.09 . Additionally, no

Table 6

Results of binomial tests of Task means (control combinations).

Control Combination	# of Positive Means	# of Negative Means	<i>p</i>
HVAC, HUB	1	5	0.109
HVAC, DASH	5	4	0.500
RADIO, HUB	0	6	0.016
RADIO, DASH	9	0	0.002

Table 7Results of Student *t*-tests (all three dependent variables).

a. RAW ERROR SCORES				
Gender	N	Mean	Std. Dev.	Std. Error
F	270	0.02489	0.18471	0.01124
M	270	-0.00700	0.13721	0.00835
Student <i>t</i> = 2.2772 <i>p</i> = 0.0232				
b. ABSOLUTE ERROR SCORES				
Gender	N	Mean	Std. Dev.	Std. Error
F	270	0.10770	0.15198	0.00925
M	270	0.10544	0.08783	0.00535
Student <i>t</i> = 0.2115 <i>p</i> = 0.8326				
c. SQUARED ERROR SCORES				
Gender	N	Mean	Std. Dev.	Std. Error
F	270	0.03461	0.24706	0.01504
M	270	0.01880	0.03195	0.00194
Student <i>t</i> = 1.0427 <i>p</i> = 0.2976				

Table 8

ANOVA summary table for absolute error scores.

Source	df	SS	MS	<i>F</i>	<i>p</i>
<u>Between</u>					
Duration (D)	2	0.0219	0.0110	*	
Task/Duration (T/D)	27	0.4099	0.0152	1.07	0.3806
<u>Within</u>					
Subject (S)	5	0.2057	0.0411	2.91	0.0159
S x D	10	0.0897	0.0090	0.63	0.7825
S x T/D	135	1.9100	0.0141		
Method (M)	2	0.1128	0.0564	**	
M x D	4	0.0261	0.0065	***	
M x T/D	54	0.9533	0.0177	1.20	0.1776
S x M	10	0.2538	0.0254	1.72	0.0752
S x M x D	20	0.3314	0.0166	1.13	0.3225
S x M x T/D	270	3.9746	0.0147		
Total	539	8.2892			

* $F'(2, 5) = 1.095, p > 0.05$ ** $F'(2, 11) = 1.992, p > 0.05$ *** $F'(4, 18) = 0.335, p > 0.05$

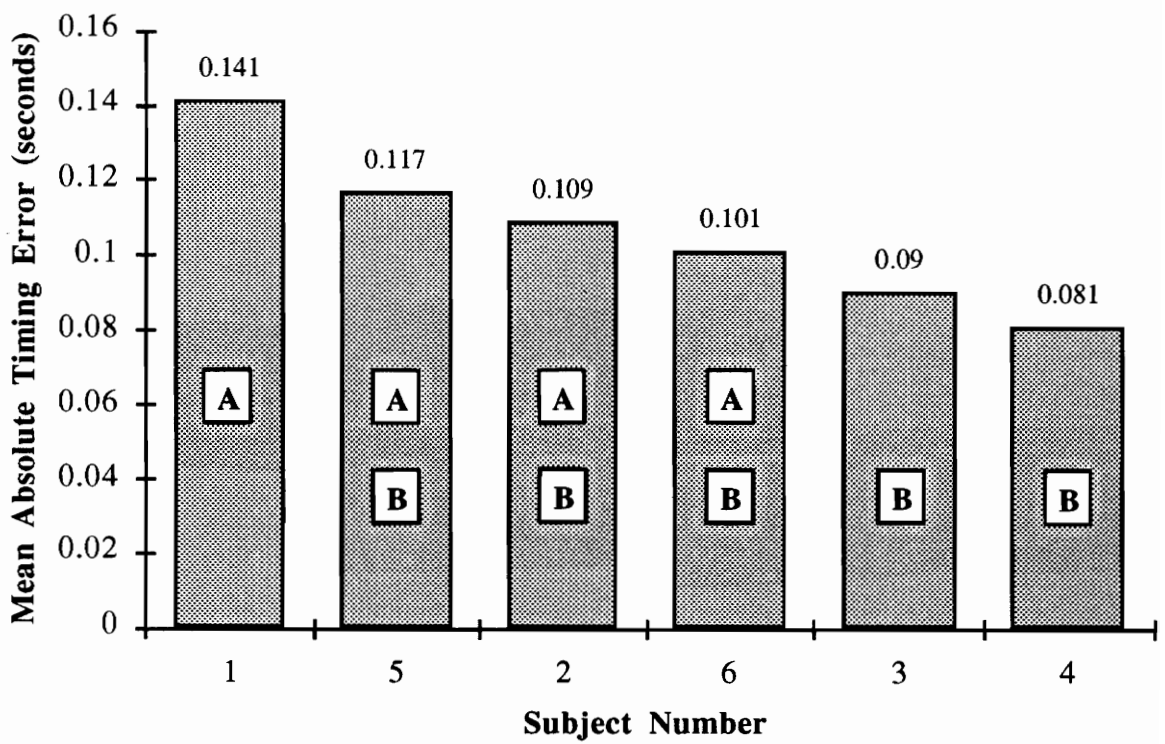


Figure 16: Mean absolute timing error as a function of Subject. Means designated by common letters do not differ significantly ($\alpha = 0.05$).

significant Gender effect was found (Table 7-c). The results of this Analysis of Variance are presented in Table 9.

Analyses of Winsorized Data

When a scatterplot of the raw error scores was created, four data points appeared as outliers. This scatterplot is shown in Figure 17, and the outlying data points are shown as solid dots. Since these outliers could most likely be attributed to nonrandom error associated with the subjects, the decision was made to perform a second complete set of analyses on a trimmed set of data.

The Winsor technique was applied to eliminate the outlying data points. Since three of the outliers were positive and only one was negative, the three highest and three lowest data points in the distribution were removed. These trimmed values were then replaced by the most extreme values remaining in each tail of the distribution. A scatterplot of the trimmed (Winsorized) data is depicted in Figure 18 .

As with the untrimmed data, three Analyses of Variance were conducted on the Winsorized data. The first ANOVA compared the ten raw performance error scores in each experimental cell to determine if there were any independent variable effects on response *bias*. The second ANOVA compared the absolute values of the ten performance error scores in each cell to investigate effects of the independent variables on response *accuracy*. Finally, the third ANOVA compared the squared values of the ten performance error scores in each cell to establish whether or not the independent variables of interest had any effect on response *consistency*. Summary tables for all three ANOVAs are presented in Appendix E, Tables E.1-E.3. Three post-hoc tests were performed along with the ANOVAS to determine if there were any significant effects of Gender. The results of these Student *t*-tests are presented in Appendix E, Table E.4.

Table 9

ANOVA summary table for squared error scores.

Source	df	SS	MS	<i>F</i>	<i>p</i>
<u>Between</u>					
Duration (D)	2	0.0288	0.0144	*	
Task/Duration (T/D)	27	0.7812	0.0289	0.95	0.5476
<u>Within</u>					
Subject (S)	5	0.2950	0.0590	1.93	0.0937
S x D	10	0.2324	0.0232	0.76	0.6673
S x T/D	135	4.1314	0.0306		
Method (M)	2	0.0150	0.0075	**	
M x D	4	0.1094	0.0273	***	
M x T/D	54	1.7536	0.0325	1.03	0.4219
S x M	10	0.2243	0.0224	0.71	0.7122
S x M x D	20	0.6609	0.0330	1.05	0.4033
S x M x T/D	270	8.4951	0.0315		
Total	539	16.7271			

* $F'(2, 5) = 0.667, p > 0.05$ ** $F'(2, 7) = 0.320, p > 0.05$ *** $F'(4, 14) = 0.803, p > 0.05$

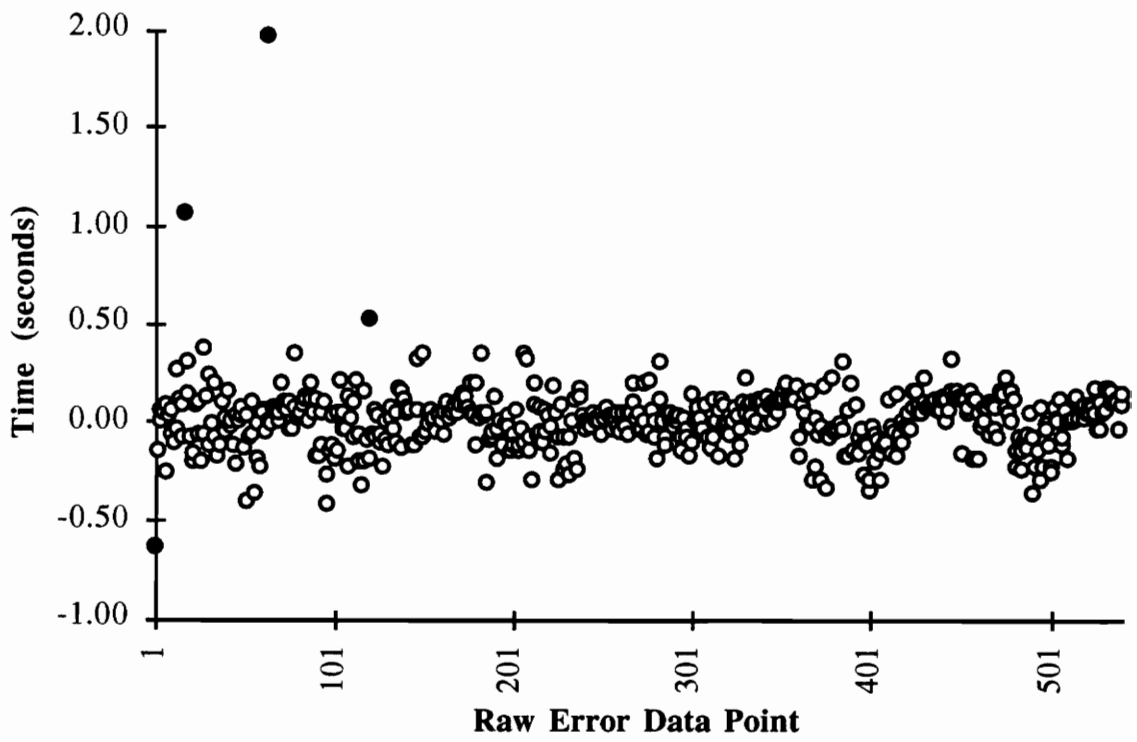


Figure 17: Scatterplot of the 540 raw error scores. Outliers are shown as solid dots.

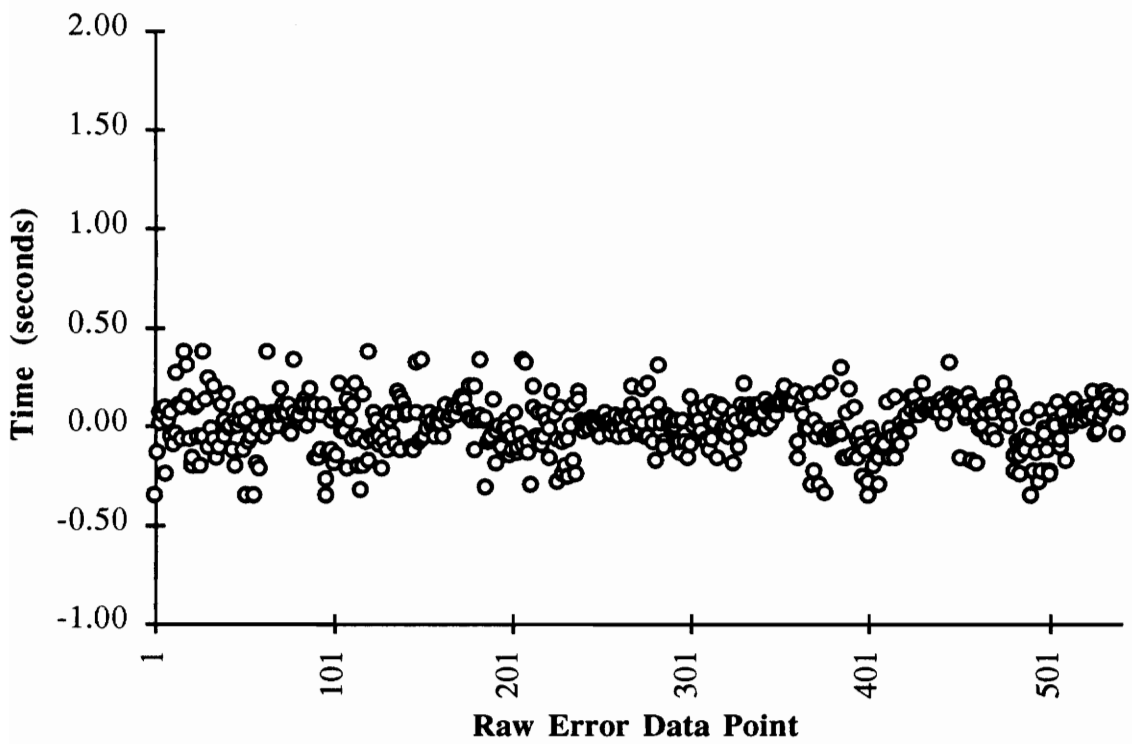


Figure 18: Winsorized scatterplot of the 540 raw error scores.

The ANOVAs and t -tests performed on the Winsorized data revealed the same significant effects as the ANOVAs and t -tests performed on the untrimmed data. Thus, the outliers did not affect the previous results. However, there was one additional significant interaction that deserves special mention. When the Winsorized raw error scores were converted to absolute error scores and analyzed with an ANOVA, the Subject x Method interaction was significant, $F(10, 270) = 3.38, p = 0.0004$. This interaction is illustrated in Figure 19. As in the case of the untrimmed data, post-hoc analysis of the Subject x Method interaction for the Winsorized data was restricted to unconfounded pairwise comparisons to limit the inflation of alpha error. Once again, the Cicchetti Table for Unconfounded Comparisons (Cicchetti, 1972) was used to adjust the Tukey Honestly Significant Difference test critical difference equation ($\alpha = 0.05$). The significant differences discovered by applying the Tukey Honestly Significant Difference test are presented in Table 10.

Determination of Limits for Subject Responses (10th to 90th Percentile)

While it is undoubtedly important to understand how the independent variables in this study affected the bias, accuracy, and consistency of subject performance, it is equally important to predict the limits within which the raw performance error scores can be expected to fall when hand-off-wheel durations are measured in an applied environment. These response limits provide valuable, additional engineering data for use in future automotive studies.

To determine the limits for subject responses, the raw error scores were first divided into three groups of data corresponding to the three variations of the human servo technique that were used in this study. For each of the three techniques, the data were grouped into four categories representing the four different control combinations used to perform a task

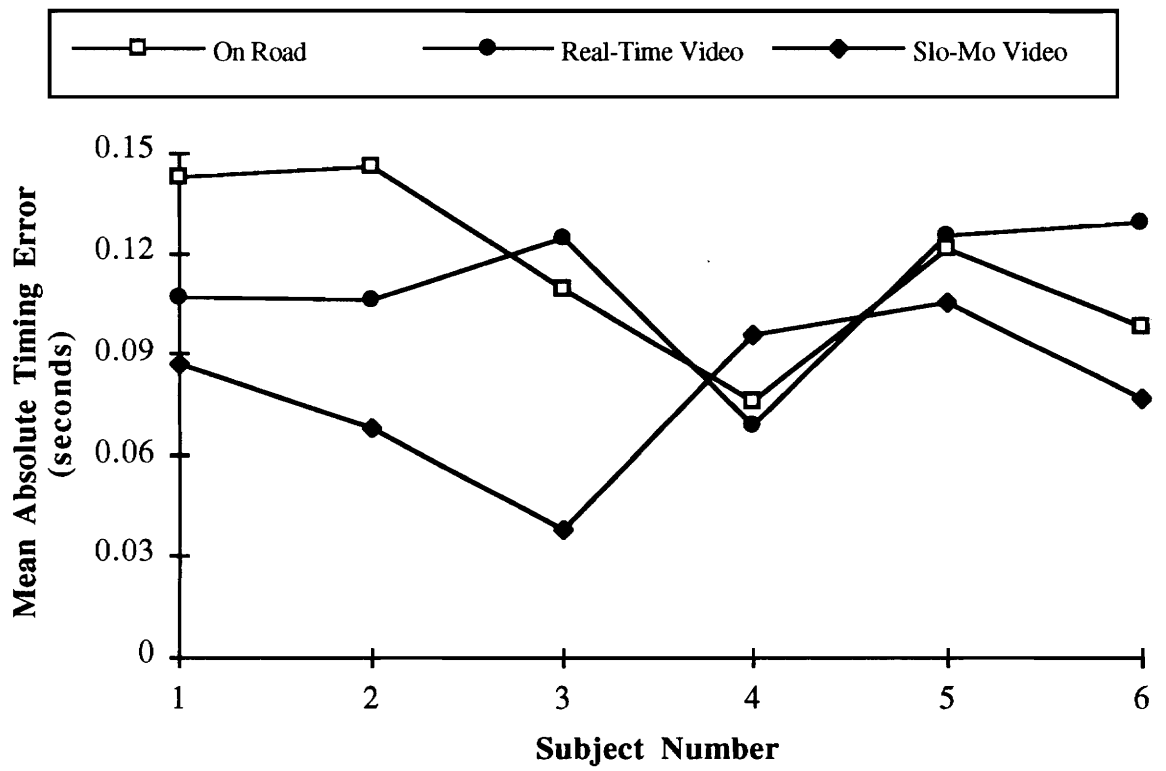


Figure 19: Subject x Method interaction (Winsorized absolute error data).

Table 10

Significant results of unconfounded pairwise comparisons for the absolute-error-scores Subject x Method interaction (Winsorized data), using the Tukey Honestly Significant Difference test.

Critical Difference _T = 0.06228 ($\alpha = 0.05$)			
Comparison		Difference Between Means	
Subject 2: Slow-Motion Video (mean = 0.06833)	vs.	On-Road (mean = 0.14633)	0.07800
Subject 3: Slow-Motion Video (mean = 0.03767)	vs.	On-Road (mean = 0.10867)	0.07100
Subject 3: Slow-Motion Video (mean = 0.03767)	vs.	Real-Time Video (mean = 0.12433)	0.08667
Slow-Motion Video: Subject 3 (mean = 0.03767)	vs.	Subject 5 (mean = 0.10467)	0.06700
On-Road: Subject 1 (mean = 0.14300)	vs.	Subject 4 (mean = 0.07633)	0.06667
On-Road: Subject 2 (mean = 0.14633)	vs.	Subject 4 (mean = 0.07633)	0.07000

(Radio/Hub, Radio/Dash, HVAC/Hub, and HVAC/Dash). For each of these four categories, the mean raw error value and the standard deviation of the raw error values were calculated. Based upon this data, the 10th percentile and 90th percentile raw error values were obtained.

The means, standard deviations, 10th percentile raw error scores, and 90th percentile raw error scores for the four types of tasks are presented in Table 11. These descriptive statistics are divided into three tables representing each of the three human servo techniques. Based upon these numbers, one can predict how observers can be expected to perform using variations of the human servo technique.

Time Requirements for Manual Timing Techniques

During the course of this study, accurate records were kept of the time required for the experimenter to use the frame-by-frame technique and for the subjects to use each of the three human servo techniques. These records reflected the amount of time involved from the beginning of the instructions for the first practice task to the end of the hand-off-wheel period for the last experimental task. The time requirements for each of the four manual timing techniques were averaged across sessions, and the following results were obtained: 38.20 minutes for the on-road servo technique, 38.96 minutes for the real-time video servo technique, 78.63 minutes for the slow-motion video servo technique, and 71.57 minutes for the frame-by-frame technique. According to these results, the frame-by-frame technique required about 87% more observation time than the on-road servo technique and approximately 84% more observation time than the real-time video servo technique. As for the slow-motion video technique, it required about 106% more observation time than the on-road servo technique and around 102% more observation time than the real-time video servo technique.

Table 11

Means, standard deviations, 10th percentile raw error scores, and 90th percentile raw error scores for four different task types (in seconds).

ON-ROAD SERVO					
Control Combination	N	Mean	Std. Dev.	10th %ile	90th %ile
RADIO, HUB	36	-0.0933	0.1360	-0.2674	0.0808
RADIO, DASH	54	0.0463	0.1211	-0.1087	0.2013
HVAC, HUB	36	-0.0389	0.2460	-0.3538	0.2760
HVAC, DASH	54	0.0417	0.1559	-0.1579	0.2413
TOTAL	180	-0.0001	0.1739	-0.2227	0.2225
REAL-TIME VIDEO SERVO					
Control Combination	N	Mean	Std. Dev.	10th %ile	90th %ile
RADIO, HUB	36	-0.1006	0.1410	-0.2811	0.0799
RADIO, DASH	54	-0.0167	0.1073	-0.1540	0.1206
HVAC, HUB	36	-0.0928	0.1193	-0.2455	0.0599
HVAC, DASH	54	-0.0241	0.1336	-0.1951	0.1469
TOTAL	180	-0.0509	0.1295	-0.2167	0.1149

Table 11 (continued)

SLOW-MOTION VIDEO SERVO					
Control Combination	N	Mean	Std. Dev.	10th %ile	90th %ile
RADIO, HUB	36	0.0561	0.0790	-0.0450	0.1572
RADIO, DASH	54	0.1089	0.2636	-0.2285	0.4463
HVAC, HUB	36	0.0514	0.0738	-0.0431	0.1459
HVAC, DASH	54	0.0787	0.0712	-0.0124	0.1698
TOTAL	180	0.0778	0.1578	-0.1242	0.2798

DISCUSSION

Interpretation of Results

According to the Analysis of Variance performed on the raw error scores, subjects made significantly higher estimates of task lengths when they used the slow-motion video servo technique than when they used either the real-time video servo technique or the on-road servo technique. In fact, the results of this study indicate that slow-motion video servo timing resulted in duration estimates that were generally greater than the true task durations, and the other two timing methods lead to duration estimates that were generally equal to or less than the true task durations. This result was unanticipated. Use of the slow-motion video servo technique was expected to result in raw error scores approximating zero, and use of the other techniques was expected to result in more positively biased error scores than those revealed from using the slow-motion video servo technique.

It is speculated that subjects using the slow-motion video servo technique took advantage of the technique's *apparent* accuracy by either (1) starting the hand-held timer at the instant they saw the driver move his hand by the slightest amount, or (2) hesitating to stop the hand-held timer until they were absolutely certain the task was complete. This would account for the positive raw error mean associated with this technique. There also appears to be a reasonable explanation for the negative raw error mean associated with using the real-time video servo technique. Simple reaction times preceding initial button-presses are believed to have been greater than simple reaction times preceding terminating button-presses. It is proposed that subjects could not time their initial button presses to coincide with the driver's hand-*off*-wheel movements as easily as they could time their ending button-presses to coincide with the hand-*on*-wheel movements, since the return of the driver's hand to the steering wheel served as a "countdown timer" for predicting when

it would make contact. As a result, task durations measured with the real-time video servo technique were generally underestimated. This same reasoning would suggest that the raw error mean for the on-road servo technique should have been more negative, because this was also a real-time measure. However, when subjects used the on-road technique, they may have taken special care to make sure their time recordings were accurate since they knew that they could not “rewind” a task like they could with the video techniques. Additionally, the fact that subjects could see the driver’s hand movements better in real life than on videotape (due to the resolution limitations of the videotape) might help to account for the minimal error scores obtained from using the on-road technique.

Although the raw error means suggest that the slow-motion video technique will lead to positive biases, the real-time video technique will lead to negative biases, and the on-road servo technique will not lead to any appreciable biases, it must be remembered that there was no significant difference between the two real-time techniques. Had a larger sample size been used, however, a significant difference between the two real-time methods might have been revealed. If this was shown to be true, then it could be said with greater confidence that the on-road technique is the most bias-free.

While it is not surprising (in light of the research hypotheses) to learn that there was no effect of Method on response *consistency*, the fact that there was no significant effect of Method on response *accuracy* was quite unexpected. As before, it was postulated that use of the slow-motion video servo technique would result in significantly smaller absolute error scores than use of the other two human servo techniques. Why there was no effect at all, either as hypothesized or as witnessed for the raw error score data, is puzzling.

As anticipated, there was a significant main effect of Subject in the ANOVA that was performed on the absolute error scores. This result demonstrates that some individuals were more accurate in timing in-car manual tasks than others. This variability between

subjects can most likely be attributed to differences in simple reaction time. In fact, it would have been surprising if all subjects had been equally accurate in their ability to measure in-car task durations. Human factors research (Boff and Lincoln, 1988; Huchingson, 1981; Johnson, 1923; Sanders and McCormick, 1987; Teichner, 1954; Wickens, 1984) points to a broad assortment of human-performance-related factors which can affect simple reaction time. Because individuals differ in terms of such variables as motivation, fatigue, and attention, to name a few, these differences can be expected to exhibit themselves through measures of human performance.

It is interesting to note that the mean absolute performance errors of subjects in this study ranged from 0.08 seconds to 0.14 seconds. These values are meaningful in light of the fact that a simple reaction to a strong, anticipated signal takes from 0.10 seconds to 0.20 seconds (Grandjean, 1986; Huchingson, 1981; Keele, 1986; Sanders and McCormick, 1987; Van Cott and Kinkade, 1972). The mean absolute performance errors demonstrate that subjects were generally able to keep the differences between their starting and ending simple reaction times below the duration of a single simple reaction. Considering the lengths of the tasks that were being timed and the inherent variability of hand-off-wheel times, such error values are probably acceptable in most analysis situations.

An unforeseen result of this study is the lack of a significant Subject effect for either the raw performance error scores or the squared performance error scores. For the same reason cited previously (simple reaction time variability), it was anticipated that the Subject effect would be significant for these two additional dependent variables. It is difficult to find an explanation for these unexpected experimental outcomes.

The raw-error-score Subject x Method interaction indicates that subjects differed in terms of (1) the extent to which the three human servo methods they used resulted in

response bias, and (2) the bias they demonstrated for a particular method. Although slow-motion video servo timing resulted in duration estimates that were *generally* greater than the true task durations and the other two timing methods lead to duration estimates that were *generally* equal to or less than the true task durations, some subjects showed more of a distinction between their performance with alternate human servo techniques than other subjects. Additionally, two subjects differed with regard to the amount of positive bias they demonstrated when using the slow-motion video servo technique.

The absolute-error-scores Subject x Method interaction demonstrates that subjects differed in terms of (1) the extent to which the three human servo methods they used affected their response accuracy, and (2) the accuracy they demonstrated for a particular method. All subjects were equally accurate in their performance with the two real-time servo techniques. However, two subjects showed significantly greater accuracy when they used the slow-motion technique than when they used the other two techniques. Furthermore, a few subjects differed with regard to their accuracy in using the on-road servo technique, and two subjects differed with regard to their accuracy in using the slow-motion video servo technique.

The fact that the ANOVA performed on the untrimmed absolute error scores did not reveal a significant Subject x Method interaction is interesting. If the *p*-value for the absolute-error-score Subject x Method interaction (Winsorized data) had been close to the established alpha level of 0.05, this interaction could have been “explained away” as being an artifact of the Winsor technique. However, since the *p*-value for the absolute-error-score Subject x Method interaction (Winsorized data) was only 0.0004, the potential significance of the interaction effect for the untrimmed data was obviously obscured by a relatively small number of outliers.

Both of the significant Subject x Method interactions were unanticipated. It was hypothesized that, even though subjects would most likely differ from one another in terms of bias, accuracy, and consistency, each would exhibit the same general performance relationships across the three human servo techniques. This hypothesis was based on the absence of empirical evidence arguing otherwise. As a result, it is surprising that the results turned out as they did.

Although the ANOVA performed on the raw error scores demonstrated a significant Task effect, this result is not necessarily favorable. As discussed earlier, the study was designed to include three different types of tasks: short duration, medium duration, and long duration tasks. Since one of the original hypotheses was that there would be a significant effect of task Duration for at least two of the three ANOVAs, it was also assumed that there would be significant differences between tasks of different durations for those same two ANOVAs. There was no significant Duration effect, however. As a result, there should have been no significant differences between individual tasks. The fact that there *was* a significant Task effect for the raw error score data suggests that there was nonequivalency of tasks; the ten tasks in each experimental cell may not have necessarily been replications of a particular treatment combination. Fortunately, a significant Task effect existed for only one of the three ANOVAs.

According to the non-parametric analyses performed on the 30 Task means (raw error scores), there were significantly more underestimations of hub-based task durations than overestimations. There were also significantly more overestimations of dash-based task durations than underestimations. Furthermore, the data indicate that subjects tended to underestimate the durations of hub-based radio tasks and overestimate the durations of dash-based radio tasks.

The discussion of foreperiods presented in the literature review suggests that the aforementioned results should have gone in the opposite direction. At the end of a hub-based task, the observer's simple reaction foreperiod will generally be shorter than the simple reaction foreperiod occurring at the end of a dash-based task. This fact is due to the shorter travel distance required for the driver's hand to move from the hub to the steering wheel. As already established, simple reaction time will decrease with foreperiod length when variable foreperiods are involved (as they are in this study). Thus, simple reaction time for the second button press on a hub-based task should be greater than that for the second button press on a dash-based task. Assuming the beginning simple reaction times for both types of tasks are similar, this would lead one to predict larger estimations of hub-based task durations than dash-based task durations. Additionally, because of such short warning intervals occurring at the ends of hub-based tasks, observers would be expected to have inadequate response preparation time, and simple reaction time should supposedly suffer. This would also lead one to predict larger estimations of hub-based task durations than dash-based task durations. Since the results obtained from the non-parametric analyses of Task means are not what the literature would predict, there must be some other explanation.

It is speculated that subjects perceived the ending foreperiods for dash-based tasks to be relatively longer than the ending foreperiods for hub-based tasks, which they were, since the travel distance from the dash controls to the steering wheel was farther than the travel distance from the hub controls to the steering wheel. If this speculation is true, they may have been "liberal" about stopping the timer at exactly the right instant when timing dash tasks, for they were not as rushed to do so. When they were timing hub-based tasks, however, they may have felt rushed, and so they stopped the timer "conservatively" (early)

to avoid making too great an error. This would explain the overestimations of dash-based task durations and the underestimations of hub-based task durations.

Another potential explanation for the subjects' underestimations of hub-based task durations has to do with expectancy. Most drivers are accustomed to using their right hand to perform in-car manual tasks. They seldom have to use their left hand to perform a secondary task. However, this experiment required the driver to use his left hand to perform certain hub-based tasks. At times, subjects may have been expecting the driver to use his right hand to perform a particular hub-based task, and so they watched that hand to see when it would begin to move. If it was actually a left-handed task that was to be performed, there would be a simple reaction time delay as the subjects refocused their attention from the driver's right hand to his left hand and then commenced timing. If simple reaction time is greater at the beginning of a timing period than at the end, the task duration will be underestimated.

The significant effect of Gender on error bias (untrimmed raw error data) was completely unanticipated. Considering that nothing in the reviewed reaction time literature suggests that females are any different from males in terms of simple reaction time, there was no reason to believe that there would be a difference between the two groups of subjects. Why females overestimated task durations and males made slight underestimations may be answered only by undertaking further research.

As stated above, and contrary to what was predicted, there was no significant effect of Duration on any of the dependent variables of interest. As for the hypothesis of no significant Subject x Duration interaction, this hypothesis was supported. There was also no statistically significant evidence of any Method x Duration interaction or Subject x Method x Duration interaction.

Analysis of the time requirements for the four manual timing techniques (frame-by-frame technique and three human servo techniques) revealed that the on-road servo technique and the real-time video servo technique each demand approximately the same amount of observation time. The slow-motion video servo technique requires approximately twice as much observation time as the on-road and real-time servo techniques, and the frame-by-frame technique requires a little less than twice as much observation time as the on-road and real-time servo techniques. It was expected that the slow-motion video servo technique and the frame-by-frame technique would demand far more observation time than they actually did. The fact that these two techniques did not require nearly as much time as expected was a welcome finding; this implies that the “slower” methods are more usable than they were presumed to be.

Lastly, in analyzing the time requirements for the frame-by-frame technique, an interesting phenomenon was identified. As the experimenter performing the frame-by-frame analyses worked through the six videotapes, his analysis time decreased. In fact, the first videotape required 96.39 minutes to analyze, and the last videotape required only 58.21 minutes to analyze. These results suggest that practice can lead to considerably reduced analysis time when using the frame-by-frame technique.

Difficulties with Experimental Apparatus and Procedures

While this study was being run, two issues relating to difficulties with the experimental apparatus and procedures arose which may (or may not) have had a considerable impact on the final results. When considering the results, one is advised to take these difficulties into account. Furthermore, if this study is to be replicated, it is highly recommended that the problems be rectified.

First, when using either of the two video servo techniques, subjects occasionally complained of trouble identifying the exact task start and stop points because of excessive washout on the video images. Varying outdoor lighting conditions were undoubtedly the cause of this problem. Fortunately, there are a few potential solutions to this predicament. For one, on-road trials can take place in the late afternoon when the sunlight is not too intense, provided that there is a lack of excessive traffic congestion on the test route. Additionally, the camera images can be periodically monitored by the timing analyst while the experiment is occurring, so that he or she can make iris adjustments at the instant a washout problem occurs. Finally, one subject suggested using a color camera to record hand movements off and onto the steering wheel rather than a monochrome camera. It is speculated that the color contrast might make it easier to distinguish the hand from its background.

Subjects also complained of trouble identifying exact task start and stop points at times when the driver's fingers were obscured by the steering wheel rim. As previously mentioned, each subject was told to start the timer at the instant the driver's hand lost contact with the steering wheel and to stop the timer at the instant the driver's hand made return contact with the steering wheel. Occasionally, though, the driver would not remove his hand in one discrete movement. Rather, he would start by lifting his fingers off the back of the steering wheel rim (which a subject could not see), and he would then lift his palm off the side of the rim. Furthermore, the driver would not always make return contact in one discrete movement. He would first hit the side of the steering wheel rim with his palm, and he would then wrap his fingers around the back of the rim. Although it would be extremely obtrusive to instruct a driver to grip the steering wheel in a particular manner, and it would be difficult to find a camera position that would alleviate the problem, three alternative solutions might serve to remedy this difficulty.

For one, a touch-sensitive strip could be attached to the steering wheel and serve as an on/off switch for an electronic timer. This technique was discussed in some detail at the beginning of this paper. Although the use of a touch-sensitive strip may alleviate the problem of identifying exact task start and stop points, it may also create new problems (such as false activations and deactivations). Clearly, this technique must be subject to further investigation to determine if it is truly more accurate than any variations of the human servo technique.

Second, it might be beneficial to redefine hand-off-wheel time. For example, rather than instructing a timing analyst to begin timing at the instant the hand loses contact with the steering wheel and to end timing at the instant the hand makes return contact with the steering wheel, the analyst can be told to begin timing when the hand *clears* the wheel and to end timing once any part of the hand *touches* the wheel. By redefining hand-off-wheel time to include easily observable start and stop points, it may be possible to eliminate some of the biases associated with using the human servo technique.

Third, and possibly most important, observers can be trained prior to measuring the durations of in-car manual tasks. This training period might even be designed to encompass an alternative operational definition of hand-off-wheel time, such as the one mentioned in the preceding paragraph. By undergoing a period of detailed instruction and practice, it is hypothesized that an observer might be able to discern task start and stop points with considerably greater ease. In addition, the negative biases associated with timing hub-based tasks and the positive biases associated with timing dash-based tasks might be reduced (or even eliminated) through proper training. In short, if the human servo technique is to be implemented in future automotive studies, it is strongly suggested that extensive timing analyst training be provided.

Conclusions

Frame-by-frame video analysis is commonly accepted as being the most precise way of measuring the durations of in-car manual tasks. Unfortunately, this technique requires a considerable investment in money, equipment, and time (although not as much as originally anticipated) to be used properly. The present experiment was primarily designed to compare the accuracy, bias, and consistency of three variations of the human servo technique, and to determine whether any of these variations approach the preciseness associated with the frame-by-frame video analysis technique.

The results from this study indicate that the three human servo techniques differ little in terms of accuracy and consistency. However, tasks timed using the slow-motion video servo technique will show positive biases (the task lengths will tend to be overestimated) in comparison to the other techniques. Since mean raw error using the on-road servo technique was the closest to zero in this experiment, it is suggested that this human servo method be used when timing in-car tasks. In addition to approaching the preciseness of the frame-by-frame video analysis technique, this method requires minimal commitment of time, money, and resources. Furthermore, results of using the on-road servo technique are instantly available for analysis; no further data reduction is required.

If an automotive study is being conducted that demands an *extremely* high level of accuracy in the measurement of in-car manual task durations, the frame-by-frame technique is still a feasible alternative to the on-road servo technique. The present study indicates that a trained observer can use the frame-by-frame technique with only a moderate commitment of time. If the equipment is available/affordable for performing task duration measurements in this manner, and if the task duration data is not needed instantly, then it is recommended that the frame-by-frame technique be used. However, it must be emphasized that this technique is *only* recommended when close to the highest level of accuracy is

demanded. Under any other conditions, the on-road servo technique is considerably more cost-effective and practical.

This experiment also reveals that observers can be expected to time in-car manual tasks with varying degrees of accuracy. This finding leads to the suggestion that only one observer be used to time all in-car manual tasks using the on-road human servo technique. By having just one individual perform all of the timing, the accuracy of the recorded times will be consistent throughout the entire study. As long as this accuracy is consistent, it should not serve as a confound for the rest of the results. It is further recommended that observers be pre-screened on the basis of simple reaction time so that accuracy can be maximized in the actual study (assuming that simple reaction time will be related to human servo error).

The significant Gender effect and the significant Subject x Method interactions also signify that only one individual should be used to time all in-car manual tasks using the on-road human servo technique. The use of only one observer will prevent between-observer variations in bias and accuracy from having an effect on the results of the study being conducted.

This research reveals that the length of a given in-car manual task has no effect on the bias (if any), accuracy, or consistency with which it is timed. Nonetheless, in-car tasks should be carefully chosen if they are to serve as replications for a particular treatment combination in an experiment. The study indicates that pairs of tasks can differ with regard to the bias with which they are timed; pre-testing of tasks may be the only way to ensure that they are fit to serve as replications. Moreover, the results from this study reveal that hub-based task durations might be underestimated and dash-based task durations might be overestimated when using a variation of the human servo technique. If timing analysts are

carefully trained prior to participating in a study, they may be able to reduce some of the biases associated with timing hub- and dash-based tasks.

Finally, this research led to the derivation of a body of additional engineering data which can be used to predict analyst error in timing various types of in-car manual tasks (Table 11). These data can also assist future experimenters in determining which variation of the human servo technique is most appropriate for a given study, as well as the costs associated with a given technique. In addition, the data on observer error variability can be used in interpreting the results of future automotive studies which involve the timing of driver tasks.

Transportation research studies are commonly known to require a large commitment of time and money. Expensive experimental apparatus, payments made for subject participation, and time involved in data analysis can all be guaranteed to consume research resources in one way or another. Anything that can be done to save researchers dollars and/or hours is usually welcome. Using the on-road servo technique to time in-car manual tasks, rather than the frame-by-frame video analysis technique, is just one way to increase the overall efficiency of transportation research studies without compromising the accuracy of the empirical data.

Suggestions for Future Research

Based on the findings of this study, several areas for future research have been identified. These potential research areas are listed below:

- (1) Analyze the accuracy of the three human servo techniques when they are applied in alternate experimental settings. Since the measurement of human behavior durations is so common in human factors research, there are likely to be

additional situations in which one (or more) of the human servo techniques might be used in place of frame-by-frame video analysis.

- (2) Determine the effects of timing analyst position on performance while using the on-road servo technique. Specifically, compare analyst performance while sitting in the front seat of a research vehicle with performance while sitting in the back seat.
- (3) Derive a set of equivalent in-car manual tasks that can be used as replications for a particular treatment combination in a related study. Data from the tasks used in this study might provide a basis for developing the new set of tasks.
- (4) Examine the differences between males and females in terms of accuracy in timing human behaviors.
- (5) Compare frame-by-frame measurement of hand-off-wheel time with computerized measurement of hand-off-wheel time using a touch-sensitive strip mounted on the steering wheel. Determine if one technique is any more accurate than the other, and if not, where the differences lie.

REFERENCES

- Argyle, M., and Graham, J.A. (1976). The Central Europe experiment: Looking at persons and looking at objects. *Environmental Psychology and Nonverbal Behavior*, 1 (1), 6-16.
- Bertelson, P.B. (1967). The time course of preparation. *Quarterly Journal of Experimental Psychology*, 19 (3), 272-279.
- Bertelson, P., and Tisseyre, F. (1968). The time-course of preparation with regular and irregular foreperiods. *Quarterly Journal of Experimental Psychology*, 20 (3), 297-300.
- Bevan, W., Hardesty, D.L., and Avant, L.L. (1965). Response latency with constant and variable interval schedules. *Perceptual and Motor Skills*, 20, 969-972.
- Boff, K.R., and Lincoln, J.E. (Eds.). (1988). *Engineering data compendium: Human perception and performance*. Dayton, OH: AAMRL, Wright-Patterson Air Force Base.
- Bradley, J.V. (1975). The optimal-pessimal paradox. *Human Factors*, 17 (4), 321-327.
- Buckolz, E., and Wilberg, R. (1975). A priori probability effects on simple reaction time. *Journal of Motor Behavior*, 7 (1), 73-80.
- Cherulnik, P.D., Neely, W.T., Flanagan, M., and Zachau, M. (1978). Social skill and visual interaction. *The Journal of Social Psychology*, 104, 263-270.
- Cicchetti, D.V. (1972). Extension of multiple-range tests to interaction tables in the analysis of variance: A rapid approximate solution. *Psychological Bulletin*, 77 (6), 405-408.
- Drazin, D.H. (1961). Effects of foreperiod, foreperiod variability, and probability of stimulus occurrence on simple reaction time. *Journal of Experimental Psychology*, 62 (1), 43-50.

- Goldberg, M.L., and Wellens, R.A. (1979). A comparison of nonverbal compensatory behaviors within direct face-to-face and television-mediated interviews. *Journal of Applied Social Psychology*, 9 (3), 250-260.
- Gottsdanker, R. (1982). Age and simple reaction time. *Journal of Gerontology*, 37 (3), 342-348.
- Grandjean, E. (1986). *Fitting the task to the man: An ergonomic approach*. London: Taylor & Francis.
- Huchingson, R.D. (1981). *New horizons for human factors in design*. New York: McGraw-Hill.
- Johansson, G., and Rumar, K. (1971). Drivers' brake reaction times. *Human Factors*, 13 (1), 23-27.
- Johnson, H.M. (1923). General review: Reaction-time measurements. *Psychological Bulletin*, 20, 562-589.
- Karlin, L. (1959). Reaction time as a function of foreperiod duration and variability. *Journal of Experimental Psychology*, 58 (2), 185-191.
- Keele, S.W. (1986). Motor control. In K.R. Boff, L. Kaufman, and J.P. Thomas (Eds.), *Handbook of perception and human performance: Cognitive processes and performance* (pp. 30:3-30:4). New York: John Wiley and Sons.
- Klemmer, E.T. (1957). Simple reaction time as a function of time uncertainty. *Journal of Experimental Psychology*, 54 (3), 195-200.
- Lindman, R. (1980). Alcohol and eye contact. *Scandinavian Journal of Psychology*, 21, 201-205.
- Maynard, H.B. (Ed.). (1963). *Industrial engineering handbook*. New York: McGraw-Hill.

- Meister, D. (1985). *Behavioral analysis and measurement methods*. New York: John Wiley and Sons.
- Mo, S.S., and George, E.J. (1977). Foreperiod effect on time estimation and simple reaction time. *Acta Psychologica*, 41, 47-59.
- Moore, T.V. (1904). A study in reaction time and movement. *The Psychological Review*, 6, (1, Whole No. 24).
- Näätänen, R. (1971). Non-aging foreperiods and simple reaction time. *Acta Psychologica*, 35, 316-327.
- Näätänen, R., Muranen, V., and Merisalo, A. (1974). Timing of expectancy peak in simple reaction time situation. *Acta Psychologica*, 38, 461-470.
- Nickerson, R.S. (1965). Response time to the second of two successive signals as a function of absolute and relative duration of intersignal interval. *Perceptual and Motor Skills*, 21, 3-10.
- Nickerson, R.S. (1967). Expectancy, waiting time and the psychological refractory period. *Acta Psychologica*, 27, 23-34.
- Nickerson, R.S., and Burnham, D.W. (1969). Response times with nonaging foreperiods. *Journal of Experimental Psychology*, 79 (3), 452-457.
- Niemi, P. (1979). Stimulus intensity effects on auditory and visual reaction processes. *Acta Psychologica*, 43, 299-312.
- Niemi, P., and Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, 89 (1), 133-162.
- O'Donnell, R.D. (1972). *Handbook of human performance measures*. Unpublished technical report, Brain Research Institute, UCLA.
- Okamoto, Y. (1980). Contextual effect in simple reaction time. *Japanese Psychological Research*, 22 (4), 219-222.

- Polzella, D.J., Ramsey, E.G., and Bower, S.M. (1989). The effects of brief variable foreperiods on simple reaction time. *Bulletin of the Psychonomic Society*, 27 (5), 467-469.
- Possamai, C., Granjon, M., Requin, J., and Reynard, G. (1973). Sequential effects related to foreperiod duration in simple reaction time. *Perceptual and Motor Skills*, 36, 1185-1186.
- Sanders, M.S., and McCormick, E.J. (1987). *Human factors in engineering and design*. New York: McGraw-Hill.
- Shinnick, M.D. (1987). *MODAPTS plus resource manual*. Blacksburg, VA: Dynamics Research Group.
- Simon, J.R., and Slaviero, D.P. (1975). Differential effects of a foreperiod countdown procedure on simple and choice reaction time. *Journal of Motor Behavior*, 7 (1), 9-14.
- Teichner, W.H. (1954). Recent studies of simple reaction time. *Psychological Bulletin*, 51 (2), 128-149.
- Telford, C.W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology*, 14 (1), 1-36.
- Van Cott, H.P., and Kinkade, R.G. (Eds.). (1972). *Human engineering guide to equipment design*. Washington, D.C.: American Institutes for Research.
- Wargo, M.J. (1967). Human operator response speed, frequency, and flexibility: A review and analysis. *Human Factors*, 9 (3), 221-238.
- Wickens, C.D. (1984). *Engineering psychology and human performance*. Glenview, IL: Scott, Foresman and Company.
- Wilkinson, R.T., and Allison, S. (1989). Age and simple reaction time: Decade differences for 5,325 subjects. *Journal of Gerontology*, 44 (2), P29-35.

- Woodrow, H. (1914). The measurement of attention. *The Psychological Monographs*, 17, (5, Whole No. 76), 1-158.
- Zahn, T.P., and Rosenthal, D. (1966). Simple reaction time as a function of the relative frequency of the preparatory interval. *Journal of Experimental Psychology*, 72 (1), 15-19.

APPENDIX A
Task Duration and Error Log Sheets

TASK DURATION LOG SHEET

Subject Number: _____ Subject I.D. _____ Date _____

TIMING TECHNIQUE: _____

Practice Trials

1 _____

2 _____

3 _____

4 _____

5 _____

TASK DURATION LOG SHEET

Subject Number: _____ Subject I.D. _____ Date _____

TIMING TECHNIQUE: _____

Actual Trials

1	_____	16	_____
2	_____	17	_____
3	_____	18	_____
4	_____	19	_____
5	_____	20	_____
6	_____	21	_____
7	_____	22	_____
8	_____	23	_____
9	_____	24	_____
10	_____	25	_____
11	_____	26	_____
12	_____	27	_____
13	_____	28	_____
14	_____	29	_____
15	_____	30	_____

ERROR LOG SHEET

Subject Number: _____ Subject I.D. _____ Date _____

TIMING TECHNIQUE: _____

Practice Trials

1 _____

2 _____

3 _____

4 _____

5 _____

ERROR LOG SHEET

Subject Number: _____ Subject I.D. _____ Date _____

TIMING TECHNIQUE: _____

Actual Trials

1 _____
2 _____
3 _____
4 _____
5 _____
6 _____
7 _____
8 _____
9 _____
10 _____
11 _____
12 _____
13 _____
14 _____
15 _____

16 _____
17 _____
18 _____
19 _____
20 _____
21 _____
22 _____
23 _____
24 _____
25 _____
26 _____
27 _____
28 _____
29 _____
30 _____

APPENDIX B
In-Car Manual Tasks

(NOTE: Parentheses surround preparatory settings)

PRACTICE TASKS

Use the *dash* controls to set the radio to preset five. (preset three)

Use the *dash* controls to tune the radio up to 97.9 FM. (87.9 FM)

Use the *dash* controls to put the HVAC system in the upper mode, then decrease the fan speed to low. (auto mode, fan at level ten)

Use the *hub* controls to put the HVAC system in the bi-level mode. (auto mode)

Use the *hub* controls to set the radio to preset two, then decrease the volume until it is no longer audible and increase the volume to a comfortable level. (preset three, volume at level five)

SHORT TASKS

Radio Tasks

Use the *dash* controls to set the radio band to AM. (FM)

Use the *dash* controls to set the radio band to preset one. (preset 3)

Use the *dash* controls to seek the next available higher frequency radio station. (preset three)

Use the *hub* controls to set the radio band to FM. (AM)

Use the *hub* controls to seek the next available lower frequency radio station. (preset three)

HVAC Tasks

Use the *dash* controls to increase the temperature by one degree. (72 degrees)

Use the *dash* controls to decrease the fan speed by one level. (fan at level three)

Use the *dash* controls to put the HVAC system in the bi-level mode. (auto mode)

Use the *hub* controls to decrease the temperature by one degree. (72 degrees)

Use the *hub* controls to increase the fan speed by one level. (fan at level one)

MEDIUM TASKS

Radio Tasks

Use the *dash* controls to tune the radio up to 1070 AM. (570 AM)

Use the *dash* controls to decrease the radio volume until it is no longer audible, then increase the radio volume to a comfortable level. (volume at level five)

Use the *dash* controls to set the radio to preset three, then tune down to 87.9 FM. (107.9 FM)

Use the *hub* controls to increase the radio volume to a comfortable level, then decrease the radio volume until it is no longer audible. (volume at level one)

Use the *hub* controls to set the radio to preset five. (preset one)

HVAC Tasks

Use the *dash* controls to increase the fan speed to high. (fan at level one)

Use the *dash* controls to decrease the temperature to 65 degrees. (72 degrees)

Use the *dash* controls to scroll through all of the HVAC modes. (auto mode)

Use the *hub* controls to decrease the fan speed to low. (fan at level ten)

Use the *hub* controls to increase the temperature to 79 degrees. (72 degrees)

LONG TASKS

Radio Tasks

Use the *dash* controls to tune the radio down to 570 AM. (1570 AM)

Use the *dash* controls to tune the radio up to 107.9 FM. (87.9 FM)

Use the *dash* controls to set the radio to preset five, then tune down to 87.9 FM. (97.9 FM)

Use the *hub* controls to set the radio to preset five, then increase the volume to a comfortable level and decrease the volume until it is no longer audible. (volume at level one, 87.9 FM)

Use the *hub* controls to seek the next available lower frequency radio station, then decrease the volume until it is no longer audible and increase the volume to a comfortable level. (volume at level five, 97.9 FM)

HVAC Tasks

Use the *dash* controls to put the HVAC system in the bi-level mode, then increase the fan speed to high. (auto mode, fan at level one)

Use the *dash* controls to decrease the temperature to 65 degrees, then increase the temperature to 72 degrees. (72 degrees)

Use the *dash* controls to determine the outside temperature, then adjust the inside temperature to equal the outside temperature and increase the fan speed to high. (72 degrees, fan at level one)

Use the *hub* controls to increase the temperature to 79 degrees, then decrease the temperature to 72 degrees. (72 degrees)

Use the *hub* controls to put the HVAC system in the lower mode, then decrease the fan speed to low. (auto mode, fan at level 10)

APPENDIX C
General Instructions for the Experiment

MANUAL MEASUREMENT TIMING STUDY
INSTRUCTIONS GIVEN TO SUBJECT
PRIOR TO OBTAINING INFORMED CONSENT

The purpose of this experiment is to compare the accuracy of three different methods for gathering data. In automotive human factors research it is often necessary to determine the length of time that the driver has a hand off the wheel while performing in-car tasks. Such tasks might include adjusting the radio or the air conditioning, for example. You will be asked to perform timing of such events, both in an actual vehicle while riding as a passenger in the back seat and observing a driver, and in the laboratory by viewing videotapes of a driver.

This study is being conducted by the Vehicle Analysis and Simulation Laboratory, Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, telephone number: (703) 231-9084. The research team will include Barry Grant, a graduate student and research assistant in Industrial and Systems Engineering, and Dr. Walter W. Wierwille, principal investigator and P.T. Norton Professor of Industrial and Systems Engineering.

In the study you will be asked to time, using an electronic stopwatch, the *hand-off-wheel* intervals of a driver, driving on U.S. Route 460, north of Blacksburg. The driver will be an experimenter, and the individual in the front passenger seat will also be an experimenter. The experimenter in the passenger's seat will issue commands to perform in-car tasks that the driver will execute. Your job will be to time and record hand-off-wheel time for each of these commands. The in-car task you will perform will take about one hour, including training.

The research vehicle used for this research will be a late model American car with instrumentation allowing the videotaping of the subject's hands. There will be additional instrumentation in the vehicle that will not be used in this experiment.

You will also be asked to perform two similar runs in the laboratory. In one run you will determine the hand-off-wheel times from a videotape that is played in real time (normal speed). In another run, you will perform the same task, but with the tape played in slow motion. The two tapes that you will view in the laboratory will be similar to the run you will experience in the vehicle. The real-time laboratory run is expected to take about one hour of your time (including training), and the slow motion laboratory run is expected to take about two hours (including training). Thus, your total participation time in this experiment will be about 4 hours. However, this is an estimate, and your actual participation time may be somewhat longer or shorter.

The three runs in which you will participate (one in the car and two in the lab) will be administered in an order that is predetermined. Thus, depending on your subject number, you will experience the three conditions in a specified order.

If during the study you feel that you cannot continue for any reason, you have the right to terminate your participation; you will be paid for your participation up to that time. This includes the right to withdraw at any time after you have read and signed the attached informed consent form. If you withdraw during the in-car session, one of the experimenters will drive you back to Blacksburg. (You should be aware that your data cannot be used for analysis if it is incomplete.)

If you have any questions about the experiment or your rights as a participant after reading the attached informed consent form, please do not hesitate to ask. We will answer your questions as openly and honestly as possible.

There are some risks inherent in this study. They are outlined in the following informed consent form.

APPENDIX D
Participant's Informed Consent Form

PARTICIPANT'S INFORMED CONSENT

1. You are being asked to volunteer to be a subject in a research project whose purpose and description are contained in the document entitled *Manual Measurement Timing Study*, which you have already read.
2. There are some risks to which you expose yourself in volunteering for this research. The risks are:
 - a. The risk of an accident normally associated with being a passenger in an automobile in light or moderate traffic, as well as on straight and curved roadways.
 - b. The slight additional risk of an accident that might possibly occur while the driver is performing instructed tasks. Past research indicates that this risk is minimal.

The following precautions will be taken during the in-car run:

- a. You will be required to wear the lap and shoulder belt restraint system anytime the car is on the road.
- b. The vehicle will be equipped with a fire extinguisher, first-aid kit, and a CB radio (or cellular phone).
- c. The front-seat experimenter will have an override brake pedal.

There are no known discomforts in this experiment.

3. The data gathered in this experiment will be treated with anonymity. Shortly after you have participated, your name will be separated from your data.
4. While there are no direct benefits to you from this research (other than payment), you may find the experiment interesting. Your participation, along with that of other volunteers, should make it possible to improve the accuracy and efficiency of data gathering in automotive human factors.
5. You should not volunteer for participation in this research if you are under 18 years old, or if you are not in good health, or if you are pregnant.
6. You should know that the principal investigator of the research project and his associates will answer any questions that you may have about this project, and you should not sign this consent form until you are satisfied that you understand all of the previous descriptions and conditions.

You should further be aware that you may contact Dr. Ernest Stout, Chairman of the University's Institutional Review Board, if you have questions or concerns about this experiment. His phone number is (703) 231-9359.

7. You should know that at any time you are free to withdraw from participation in this research program without penalty.
8. You will be paid at a rate of \$5.00 per hour for the time you actually spend. Payment will be made shortly after you have finished your participation.
9. Signature of the volunteer and date:

I have read and understand the scope of this research project and I have no other questions. I hereby give my consent to participate, but I understand that I may stop participation if I choose to do so.

Signature _____

Date _____

10. Signature of a member of the research team and date:

Signature _____

Date _____

11. Signature of a witness, not a member of the research team and date:

Signature _____

Date _____

APPENDIX E
Analyses of Winsorized Data

Table E.1

ANOVA summary table for Winsorized raw error scores.

Source	df	SS	MS	<i>F</i>	<i>p</i>
<u>Between</u>					
Duration (D)	2	0.1548	0.0774	*	
Task/Duration (T/D)	27	1.2255	0.0454	2.73	0.0001
<u>Within</u>					
Subject (S)	5	0.1207	0.0241	1.45	0.2099
S x D	10	0.3460	0.0346	2.08	0.0300
S x T/D	135	2.2445	0.0166		
Method (M)	2	1.3039	0.6519	**	
M x D	4	0.0228	0.0057	***	
M x T/D	54	0.6561	0.0121	1.39	0.0475
S x M	10	0.5103	0.0510	5.85	0.0001
S x M x D	20	0.2141	0.0107	1.23	0.2320
S x M x T/D	270	2.3569	0.0087		
Total	539	9.1556			

* $F'(2, 20) = 1.222, p > 0.05$ ** $F'(2, 11) = 11.974, p < 0.01$ *** $F'(4, 22) = 0.403, p > 0.05$

Table E.2

ANOVA summary table for Winsorized absolute error scores.

Source	df	SS	MS	<i>F</i>	<i>p</i>
<u>Between</u>					
Duration (D)	2	0.0541	0.0270	*	
Task/Duration (T/D)	27	0.2604	0.0096	1.60	0.0433
<u>Within</u>					
Subject (S)	5	0.0849	0.0170	2.81	0.0189
S x D	10	0.0624	0.0062	1.03	0.4178
S x T/D	135	0.8145	0.0060		
Method (M)	2	0.1439	0.0720	**	
M x D	4	0.0455	0.0114	***	
M x T/D	54	0.3607	0.0067	1.23	0.1507
S x M	10	0.1844	0.0184	3.38	0.0004
S x M x D	20	0.1446	0.0072	1.33	0.1612
S x M x T/D	270	1.4711	0.0054		
Total	539	3.6265			

* $F'(2, 12) = 2.744, p > 0.05$ ** $F'(2, 11) = 3.659, p > 0.05$ *** $F'(4, 20) = 1.346, p > 0.05$

Table E.3

ANOVA summary table for Winsorized squared error scores.

Source	df	SS	MS	<i>F</i>	<i>p</i>
<u>Between</u>					
Duration (D)	2	0.0029	0.0014	*	
Task/Duration (T/D)	27	0.0255	0.0009	1.27	0.1850
<u>Within</u>					
Subject (S)	5	0.0084	0.0017	2.26	0.0518
S x D	10	0.0066	0.0006	0.89	0.5446
S x T/D	135	0.1001	0.0007		
Method (M)	2	0.0138	0.0069	**	
M x D	4	0.0010	0.0002	***	
M x T/D	54	0.0350	0.0006	1.05	0.3850
S x M	10	0.0124	0.0012	2.01	0.0325
S x M x D	20	0.0159	0.0007	1.29	0.1829
S x M x T/D	270	0.1661	0.0006		
Total	539	0.3877			

* $F'(2, 9) = 1.668, p > 0.05$ ** $F'(2, 9) = 5.453, 0.05 > p > 0.01$ *** $F'(4, 16) = 0.292, p > 0.05$

Table E.4Results of Student *t*-tests for Winsorized data.

RAW ERROR SCORES				
Gender	N	Mean	Std. Dev.	Std. Error
F	270	0.01756	0.12495	0.00760
M	270	-0.00737	0.13458	0.00819

Student *t* = 2.2303
p = 0.0261

ABSOLUTE ERROR SCORES				
Gender	N	Mean	Std. Dev.	Std. Error
F	270	0.09800	0.07926	0.00482
M	270	0.10463	0.08472	0.00516

Student *t* = -0.9390
p = 0.3482

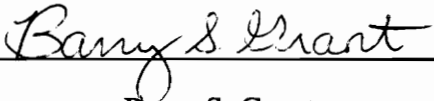
SQUARED ERROR SCORES				
Gender	N	Mean	Std. Dev.	Std. Error
F	270	0.01586	0.02586	0.00157
M	270	0.01810	0.02775	0.00169

Student *t* = -0.9683
p = 0.3333

VITA

Barry Steven Grant

Barry Steven Grant was born in West Palm Beach, Florida on December 7, 1968. He graduated from the University of Miami Honors Program with a Bachelor of Science degree in Psychology in May of 1990. Barry is currently studying Human Factors Engineering within the Industrial and Systems Engineering Department at Virginia Polytechnic Institute and State University. After graduating with a Master of Science degree, he hopes to work in either the aerospace industry or the automobile industry investigating issues related to human-machine interaction, workplace layout, control and display design, and/or vehicle simulation. Barry intends to eventually return to Virginia Tech to pursue a Ph.D. degree in Human Factors Engineering.


Barry S. Grant