



U.S. Department
of Transportation
**National Highway
Traffic Safety
Administration**



DOT HS 811 782

May 2013

Support for NHTSA's Development Of Guidelines on Distraction- Potential From Visual-Manual Interfaces

An Examination of the Definition of "Task"
And Task Taxonomies Based On Interviews
With Experts

DISCLAIMER

This publication is distributed by the U.S. Department of Transportation, National Highway Traffic Safety Administration, in the interest of information exchange. The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Department of Transportation or the National Highway Traffic Safety Administration. The United States Government assumes no liability for its contents or use thereof. If trade names, manufacturers' names, or specific products are mentioned, it is because they are considered essential to the object of the publication and should not be construed as an endorsement. The United States Government does not endorse products or manufacturers.

Suggested APA Format Citation:

Angell, L., Cook, J., & Perez, M. (2013, May). *Support for NHTSA's development of guidelines on distraction-potential from visual-manual interfaces: an examination of the definition of "task" and task taxonomies based on interviews with experts*. (Report No. DOT HS 811 782). Washington, DC: National Highway Traffic Safety Administration.

REPORT DOCUMENTATION PAGE			<i>Form Approved</i> OMB No. 0704-0188	
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE May 2013	3. REPORT TYPE AND DATES COVERED	
4. TITLE AND SUBTITLE Support for NHTSA's Development of Guidelines on Distraction-Potential From Visual-Manual Interfaces: An Examination of the Definition of "Task" and Task Taxonomies Based on Interviews With Experts			5. FUNDING NUMBERS	
6. AUTHOR(S) Linda Angell, Julie Cook, and Miguel Perez				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Virginia Tech Transportation Institute 3500 Transportation Research Plaza (0536) Blacksburg, VA 24061			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Department of Transportation National Highway Traffic Safety Administration 1200 New Jersey Avenue SE. Washington, DC 20590			10. SPONSORING/MONITORING AGENCY REPORT NUMBER DOT HS 811 782	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Document is available to the public from the National Technical Information Service www.ntis.gov			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) An interview study was conducted to gather input for the National Highway Traffic Safety Administration as it develops guidelines for preventing distraction from tasks performed with in-vehicle electronics devices. Nine experts were interviewed about the definition of "task" used in guidelines issued by the Alliance of Automobile Manufacturers (2006), along with a broader set of supporting terms used in conjunction with the definition for "task," when devices are tested for adherence to limits on distraction. These interviews explored four topics: (1) Definitions of "task" in use, (2) Other definitions used in conjunction with the definition of "task" (e.g., "goal," "subgoal," "subtask," "start state," "end state"), (3) Alternate definitions for "task," and (4) Task taxonomies. The qualitative interview data were scored for main ideas and themes. Shared ideas (versus singleton ideas) were identified through frequency counts. Among the most salient findings was that the current definition of "task" seems well accepted, at least the main portions of it. This definition gives rise to fairly common practice among users. Ideas for improving the definition and its corollary definitions were gathered. Preliminary recommendations were offered for consideration as NHTSA develops an approach to defining tasks in its guidelines to prevent and limit distraction.				
14. SUBJECT TERMS Driver distraction, secondary tasks, in-vehicle tasks, task analysis, task definition, goal, subgoal, subtask, start state, end state, task measurement			15. NUMBER OF PAGES 86	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT	

NSN 7540-01-280-5500

Prescribed by ANSI Std. Z39-18, 298-102

Standard Form 298 (rev. 2-89)

CHAPTER 1. EXECUTIVE SUMMARY

This report describes the results of an interview study gathering input for efforts currently underway at the National Highway Traffic Safety Administration to develop guidelines for preventing and limiting the distraction potential that may arise from electronic devices (such as wireless phones and route navigation systems) that have visual-manual interfaces that are integrated into vehicles, i.e., devices installed in vehicle at time of purchase. The guidelines will address in-vehicle electronic devices that allow drivers to perform tasks when the vehicle is in motion. Since the NHTSA Driver Distraction Guidelines will propose testing methodologies for determining whether a task is acceptable for performance by a driver while driving, a key issue is how a “task” should be defined in this context. To provide NHTSA with input on this issue, the interview study examined the definitions of “task” and a set of related terms and definitions.

This study focused on driver operation of electronic devices for performing non-driving-related activities while driving. Tasks are the unit of interaction between a driver and an electronic device. The tasks examined included all non-driving-related activities performed using electronic devices. Excluded from this work are such activities as driver interactions with driving controls, driver safety warning systems, any other electronic device that is necessary to drive a motor vehicle, or any other electronic device with a driver interface specified by a Federal Motor Vehicle Safety Standard.

Nine experts were interviewed about a definition of “task” currently used in the 2006 version of the *Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems* (Alliance Driver Focus-Telematics Working Group, 2006) that was developed by the Alliance of Automobile Manufacturers (hereafter, “the Alliance Guidelines”) — along with a broader set of definitions for supporting terms that are used in conjunction with the definition for “task”— when devices are tested for their adherence to limits on distraction. The Alliance Guidelines define “task” as:

“A sequence of control operations (i.e., a specific method) leading to a goal at which the driver will normally persist until the goal is reached.”

The objective of these interviews was to explore three major topic areas:

1. Definitions of “task” currently in use,
2. Other definitions of related terms that are used in conjunction with the definition of “task” (e.g., “goal,” “subgoal,” “subtask), and
3. Alternate definitions for “task.”

The experts were also interviewed about task taxonomies. Task taxonomies are categorizing tasks according to how distracting they are expected to be when performed while driving. This categorization would be based upon either easily identifiable characteristics of tasks or expert opinions as to the distraction potential of a task. (Hypothetical example: All tasks involving driver selection of one choice from a list of choices might be expected to be highly distracting.)

A task taxonomy could provide a basis on which to identify those tasks within a new information system that should be empirically tested for adherence to guidelines. In other words, if a task taxonomy could be successfully developed and implemented, then the testing called for by the

NHTSA Distraction Guidelines to determine whether a task is acceptable for performance by a driver while driving would only need to be performed for the more distracting categories of tasks. If a task taxonomy cannot be developed, then organizations would each need to determine this in their own way (e.g., through task analysis, modeling, or engineering judgment).

Results of qualitative data scoring for key themes and main ideas revealed the following:

- **Current Task Definition.** In the Alliance Guidelines, “task” is defined as “a sequence of control operations (i.e., a specific method) leading to a goal at which the driver will normally persist until the goal is reached.” This definition of “task” seems widely accepted. The definition gives rise to fairly common practice as evidenced both by what experts said and by what they did on a set of exercises in which they applied the definition to identifying the “start” and “end” points of tasks. One caveat that should be considered is that the exercises may have elicited an effect in which the experts strove to demonstrate their knowledge as opposed to what their organizations actually practice.
- **Alternative Definitions for “Task.”** There are not many already developed alternative definitions for “task.”
- **Possible Improvements to the Current Task Definition.** There are areas to consider for improving the current definition of task. They are: (a) whether the task should include leading glances or leading hand movements that occur prior to the first control operation in the task, (b) whether to include trailing glances (which occur after the last button press and may be associated with confirming that the task is finished), and (c) whether to explicitly incorporate (rather than just allow) the use of verbal indicators that identify task “start” and “end” states (e.g., the experimenter’s request to “please begin now” and the test participant’s response of “done” at completion).
- **Terms Which Support (are used in conjunction with) the Task Definition.** The interviews revealed that the set of supporting definitions of system-related terms (e.g., goal, subgoal, subtask, dependent task, transitions, and system function) is:
 - Not as widely understood (the experts working outside of the automotive companies were much less familiar with these definitions in the Alliance document than were the automotive experts).
 - Not as widely used as the task definition (perhaps reflecting disagreements or a different range of practices or needs across the sample, which included both researchers and those at automotive companies involved in product testing).
- **Taxonomies for Tasks/Test Results.** The interviews showed that:
 - Task taxonomies are not in wide use as part of distraction-testing programs (only one automotive company reported using a taxonomy as a routine part of its practices).
 - Two types of hypothetical task taxonomies shown to experts were not rated highly useful.
 - It may be too soon to propose a serious application of task taxonomies within the domain of distraction testing. Task taxonomies may require further development in order to be accepted by the community of intended users.

The findings from the interviews with a sample of nine experts led to the following preliminary recommendations:

- Adopt the term “testable task” instead of “task” in the NHTSA Driver Distraction Guidelines. Using testable task will emphasize that the NHTSA Guidelines are focused on tasks that can be tested. It will alleviate the need to encompass all possible secondary tasks that may be performed during driving. Focusing on testing will also facilitate providing greater precision in the definition of a testable task.
- Consider structuring the definition of “task” into two parts: a “core definition” (corresponding to the definitions contained in the Alliance Guidelines) and “supplemental information” (such as additional explanatory information and examples of testable tasks).
- Retain the current definition of “task” that is used in the Alliance Guidelines or slightly adapt it (following the suggestion of one interviewee):
 - A task consists of a GOAL and a specific METHOD for achieving it.
- To provide greater clarity, enhance explanatory and supplemental material (such as specific examples of testable tasks) that accompany the definition of a task.
- Retain (from the current supporting set of definitions from the Alliance document) some selected key definitions (those for “start” and “end” state are particularly important) but improve them in areas noted in this report, using examples to make the concepts concrete. This material could be made part of explanatory or supplemental material that accompanies the core definition of a testable task.
- Consider implementing a check-and-balance system to cultivate the commonality with which the definition of “task” is applied.
- Address explicitly the “scope” question (what types of devices and tasks are in scope) for guidelines and testing.
- Delay the notion of using a taxonomy for now, instead taking the time to develop the concept of a task taxonomy more fully (so it can be re-introduced later).
- Address explicitly what the expectations are about selecting tasks for guideline testing). Do all companies need to use the same method? Or do they simply need to make sure that all tasks that are within scope adhere to the guidelines?
- Address other issues that are “elephants in the room,” e.g., how to evaluate:
 - Tasks that are subject to interruptions, and
 - Task concatenations (the stringing together of tasks by a driver, one-after-another).

These sessions with experts yielded key themes and main ideas that were useful in assessing the definition of “task” currently used in the Alliance document, exploring alternatives, and suggesting avenues for possible improvement. These key themes and main ideas may provide assistance as NHTSA moves forward to develop its guidelines and as it chooses how to define and treat “tasks” within those guidelines.

GLOSSARY OF TERMS AND ACRONYMS

Alliance	Alliance of Automobile Manufacturers. A trade association comprised of automotive companies selling vehicles in the United States.
Alliance Guidelines	The Alliance has a working group (Driver Focus-Telematics Working Group [DFT]), which issued a set of voluntary guidelines in 2001 for industry to prevent or minimize driver distraction. Several subsequent updates have been made to the guidelines, the most recent in 2006. This is the <i>Statement of Principles, Criteria, and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems</i> that was developed by the Alliance of Automobile Manufacturers.
COTR	Contracting Officer's Technical Representative
ESoP	European Commission Statement of Principles
HCI	human-computer interaction
HMI	human-machine interface
HTA	hierarchical task analysis
ISO	International Organization for Standardization
JAMA	Japanese Automobile Manufacturers Association
NCAP	New Car Assessment Program
OEM	original equipment manufacturer
SAE	Society of Automotive Engineers
TOM	task order manager
TSOT	Total Shutter Open Time, a metric yielded by the occlusion test method
UL	Underwriter's Laboratory

TABLE OF CONTENTS

CHAPTER 1. EXECUTIVE SUMMARY	ii
GLOSSARY OF TERMS AND ACRONYMS.....	v
CHAPTER 1. INTRODUCTION	1
BACKGROUND	1
SCOPE	1
CHAPTER 2. METHODOLOGY.....	7
OVERVIEW OF INTERVIEWS	7
SAMPLE OF EXPERTS INTERVIEWED	8
CONFIDENTIALITY AND COMPENSATION	8
PREPARATION OF THE INTERVIEW SCRIPT AND ADVANCE PACKET	9
ADMINISTRATION OF THE INTERVIEWS.....	9
EXTRACTION OF IDEAS AND CONCEPTS FROM THE INTERVIEWS	9
ANALYSIS PROCEDURES.....	10
CHAPTER 3. RESULTS AND DISCUSSION FOR “TASK” DEFINITION	11
KEY IDEAS AND THEMES.....	11
<i>Definitions for “Task” Currently in Use.....</i>	<i>12</i>
Theme 1: The task definition in the Alliance Guidelines is satisfactory, usable, “good enough”; a “consensus” definition.....	12
Theme 2: There are no better definitions of “task” readily available.	12
Theme 3: The Alliance definition can be improved upon.	17
Theme 4: Differences in interpretation/application may be smaller than thought.....	18
Theme 5: Only one alternative definition rates similar to or better than the current one.....	30
Theme 6: Several sources of variation can give rise to differences in definition use.	32
Theme 7: Actions can be taken to minimize some of these differences in definition use.	33
<i>Concluding Comments</i>	<i>34</i>
CHAPTER 4. RESULTS AND DISCUSSION FOR OTHER SUPPORTING DEFINITIONS USED IN CONJUNCTION WITH “TASK”	35
Theme 1: The supporting definitions are “okay” but not used as widely as expected.	35
Theme 2: The set of supporting definitions can be improved.....	37
<i>Special Authors’ Note [abstracted from interviewer comments].....</i>	<i>37</i>
<i>Results Related to Specific Terms Among the Supporting Definitions</i>	<i>38</i>
Dependent Task Versus Subtask.....	38
Start State	39
End State.....	39
<i>The Reasoning behind the Use of System Function Screen</i>	<i>40</i>
CHAPTER 5. RESULTS AND DISCUSSION ABOUT TASK TAXONOMIES.....	41
HIGH-LEVEL FINDINGS ABOUT TAXONOMIES	41
Theme for Task Taxonomies: USE THEM ONLY WITH CAUTION.	41
RATINGS FOR TWO TAXONOMY TYPES.....	42
SEPARATE ISSUE: WHICH TASKS MUST BE TESTED.....	43
OTHER ISSUES MENTIONED AT CLOSE OF INTERVIEW.....	43
CHAPTER 6. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS.....	45

SUMMARY OF FINDINGS AND THEIR IMPLICATIONS	45
PRELIMINARY RECOMMENDATIONS.....	46
ACKNOWLEDGEMENTS	48
REFERENCES.....	49
APPENDIX A. INTRODUCTORY E-MAIL FOR INITIAL CONTACT WITH EXPERTS.....	50
APPENDIX B. SCRIPT FOR TELEPHONE INTERVIEWS WITH EXPERTS	52
APPENDIX C. INTERVIEW PREPARATION PACKET	64

LIST OF FIGURES

Figure 1. Overview of Interview Process	7
Figure 2. Structure of Typical “Task” Definition From Cognitive Task Analysis	16
Figure 3. Task Analysis for Destination Entry With Multiple Subtasks	21
Figure 4. Task Analysis for Manual Phone Dialing.....	22
Figure 5. Behavioral Record for Entertainment Task	23
Figure 6. Behavioral Record for Address Book Task.....	24
Figure 7. Glance Rate From Naturalistic Usage Study at VTTI During Pre-Task, Task (Event), and Post-Task Periods	28
Figure 8. Glance Durations for the Pre-Task, Task, and Post-Task Periods of Several Task Types.....	28

LIST OF TABLES

Table 1. Definitions From the Alliance Driver Focus-Telematics Guidelines..... 3

Table 2. Sample of Interviewees 8

Table 3. Positive and Negative Comments From Experts About Current Task Definition 14

Table 4. Results of Experts’ Use of Task Definitions for Actual Tasks 20

Table 5. Alternative Task Definitions That Were Rated by Experts 30

Table 6. Number of Experts Rating Each Alternative Definition Worse, Same, Better 32

Table 7. Definitions From the Alliance Driver Focus-Telematics Guidelines..... 55

Table 8. Definitions From the Alliance Driver Focus-Telematics Guidelines..... 69

CHAPTER 1. INTRODUCTION

BACKGROUND

The National Highway Traffic Safety Administration is developing guidelines to limit the distractions that may arise from electronic devices such as infotainment and route navigation systems that have visual-manual interfaces and are integrated into the vehicle by the vehicle's original manufacturer. The guidelines will address in-vehicle devices that allow drivers to perform secondary tasks when the vehicle is in motion. (Secondary tasks are defined as any interaction a driver has with an in-vehicle device that is not directly related to the primary task of driving. These tasks may relate to driver comfort, convenience, communications, entertainment, information gain, or navigation.)

Since the NHTSA Distraction Guidelines will propose testing methodologies for determining whether a task is acceptable for performance by a driver while driving, a key issue for purposes of guideline development is how a "task" should be defined.

One definition of a task is contained in the 2006 version of the *Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems* that was developed by the Alliance of Automobile Manufacturers (the Alliance Guidelines). The Alliance Guidelines define "task" as:

"A sequence of control operations (i.e., a specific method) leading to a goal at which the driver will normally persist until the goal is reached."

Differing definitions of a task can lead to differing decisions as to the actions that a driver may safely perform while driving. For example, NHTSA believes that the reason some automobile manufacturers think that route navigation system destination entry by address is acceptable under the Alliance Guidelines while other automobile manufacturers consider it to be unacceptable under the same guidelines is due to the companies differing interpretations of the precise meaning of the definition of a "task." To provide NHTSA with input on this issue, the interview study reported here was undertaken to examine definitions of "task" currently in use and a set of related issues.

This study focused upon the driver operation of electronic devices for performing non-driving-related activities while driving. Tasks are the unit of interaction between a driver and an electronic device. The tasks examined include all non-driving-related activities performed using electronic devices. Excluded from this work are driver interactions with driving controls, driver safety warning systems, any other electronic device that is necessary to drive a motor vehicle, or any other electronic device that has a driver interface that is specified by a Federal Motor Vehicle Safety Standard.

SCOPE

The main objective of this work was to provide input and recommendations about the **definition of a task** for consideration by NHTSA in preparation of its guidelines for visual-manual interfaces. The notion of "task" in this context refers to a *secondary task* (operating a radio or CD player, using a route navigation system, talking/listening/dialing a wireless phone, etc.),

which is typically performed concurrently with the primary driving tasks (e.g., controlling lane keeping, speed, headway; monitoring the forward road scene for objects and obstacles in order to maintain safety; planning and making maneuvers that may be needed to accomplish the mission of the trip). Secondary tasks, as a class, are thus distinct from primary driving tasks and are *discretionary* in nature (i.e., they may be performed when the driver chooses and when roadway demands allow). In addition, whereas many primary driving tasks are continuous in nature while the vehicle is in motion (such as the task of controlling speed), most secondary tasks are not continuous (but are typically performed in discrete epochs that occur during measurable periods of time).

In addressing the objective of providing input and recommendations to NHTSA, there were several questions to which NHTSA sought answers:

- Can the definition of a “task” that is used in the Alliance Guidelines be strengthened or improved?
- What issues might have been considered in arriving at this definition, and what rationale led to its wording (and to the elimination of other wordings)?
- Are there differences in how the Alliance Guidelines definition of a “task” is interpreted and applied? If so, what can be done to standardize practice?
- What can be done to assure *future technologies* can be accommodated by these guidelines?

A task taxonomy could provide a basis on which to identify those tasks within a new information system that should be empirically tested for adherence to guidelines. In other words, if a task taxonomy could be successfully developed and implemented, then the testing called for by the NHTSA Distraction Guidelines to determine whether a task is acceptable for performance by a driver while driving would only need to be performed for the more distracting categories of tasks. If a task taxonomy cannot be developed, then organizations would each need to determine this in their own way (e.g., through task analysis, modeling, or engineering judgment).

A secondary objective of this work was to address the following question: What approach should be taken to develop **taxonomies** that categorize the large array of secondary tasks that are and will be available to the driver? In other words:

- What would an initial taxonomy look like based on distraction potential?
- What would its implications be for testing?
- Can a task taxonomy comprehend and help with future technology?

To meet these two objectives, a group of subject matter experts were interviewed individually on these topics and issues. The following activities were conducted to prepare for the interview and to analyze the interview data.

- Review material of relevance
 - Definitions from the Alliance Guidelines and development of those definitions
 - Literature (to identify alternative definitions and new, relevant findings)
- Create an interview guide
 - Identify reasons for any possible misinterpretation or misapplication of Alliance Guidelines’ definitions
 - Propose improvements in definitions
 - Identify potential task taxonomies of use

- Seek input from a sample of experts through telephone interviews
 - Synthesize all input, recommendations, and improvements based on interview data and its analysis

The review of material revealed that the voluntary guidelines issued by the Alliance of Automobile Manufacturers Driver-Focus Working Group (2006) provide both a definition of “task” and a hierarchical set of systems definitions to operationalize testing performed to evaluate distraction.

Table 1 shows the entire set of definitions (including the definition of “task”) in the Alliance Guidelines.

Table 1. Definitions From the Alliance Driver Focus-Telematics Guidelines¹

Term	Definition	Special Notes
System Function	A system function consists of a major grouping of related tasks and is defined to be a broad category of activity performed by a system (for example, navigation). Other examples would be telecommunication-based services, Internet services, etc. (p.64)	
Goal	A goal is defined as a system state sought by a driver. Driver goals can be met through different system executions and, as such, the system states that correspond to the meeting of these driver goals can be observed and recognized (regardless of the system being operated). That is, goal achievement is defined as equivalent to achieving the system state that meets the driver's intended state, independent of the particular system being executed or method of execution. Examples given (p.64)	<p>NOTE 1: It may be necessary to operationalize a task's end state for evaluation purposes (see “End State” definition).</p> <p>NOTE 2: Clarification regarding multiple ways to reach a goal: When a system provides redundant controls or multiple software-driven paths for a user to reach a goal, all design-intended paths to reach a goal should meet the principles and criteria with representative, typical tasks. Examples given (p. 64)</p>
Subgoal	A subgoal is an <i>intermediate</i> state on the path to the goal toward which a driver is working. It is often distinguishable from a goal in two ways: (1) It is usually not a state at which the driver would be satisfied stopping, and (2) It may vary in its characteristics and/or ordering with other subgoals across hardware/interface executions and thus is system-dependent .	
Task	A task is defined as a sequence of control operations (i.e., a specific method) leading to a goal at which the driver will normally persist until the goal is reached. An example is obtaining guidance from a navigation system by entering a street address using the scrolling list method until route guidance is initiated.	NOTE: this definition only applies to secondary tasks not to the primary driving task.
Subtask	A subtask is defined as a sub-sequence of control operations that is part of a larger task sequence and which leads to a subgoal that represents an <i>intermediate</i> state in the path to	

¹ Alliance Driver Focus-Telematics Working Group (2006). Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems. Washington, D.C.: Alliance of Automobile Manufacturers. Page numbers refer to the location of these definitions and examples in the Alliance Guidelines.

	the larger goal toward which the driver is working.	
Dependent Task	<p>There is a class of tasks (called “dependent tasks”) that may be distinguished from subtasks yet cannot be initiated until another task is first completed. Their “start state” is thus dependent upon the “end state” of another (antecedent) task. However, such tasks are to be treated as tasks (rather than as subtasks) for purposes of evaluating compliance of tasks with the principles and criteria below. They can be distinguished from subtasks by examining their end state (or goal state), which will usually be a driver-sought, system-independent state. Example: After choosing a restaurant from a POI list, the driver is offered an internet function option of making a reservation at the restaurant. The dependent task of making a reservation can be initiated only following the task of selecting a restaurant within the NAV function. It is therefore a separate, dependent task (p. 65)</p>	<p>NOTE 1: Subtasks should not be treated as separate, dependent tasks. For example, entering the street name as part of the navigation destination entry is not a separate task from entering the street number; rather, these are subtasks of the same task.</p> <p>NOTE 2: The concept of “dependent tasks,” however, does have special relevance for some domains such as that of route following using a route guidance support system. In particular, after the way-finding mode has been initiated (and destination entered), subsequent route-following guidance can be treated as a series of dependent tasks. For example, following the guidance from point of issuance (“Next turn is on Maple, ¼ mile ahead”)* through achievement of goal (e.g., making of the instructed turn) would be defined as a dependent task whose start state depends on having completed the prior route maneuver successfully. (Such tasks may be evaluated analytically or through engineering judgment, as appropriate.) (p. 65) *Example added for clarity.</p>
Start State of Task	<p>The start of a task under test is the default start screen for the system function under which the task falls. Every task within a system function must share the same start state for purposes of evaluation for compliance with these principles and criteria. An exception is made for tasks that can be initiated only following the completion of a previous task. For these dependent tasks, the start screen would be the end of the previous task (p. 65).</p>	
End State of Task	<p>For the purpose of testing to the criteria contained in this section, the end state of a task is the completion of the final manual input to achieve the driver’s goal, or as indicated by the test subject, as appropriate to accurately measure the duration of the task. This operational definition of task end state is necessary due to the fact that test systems may need to be used for evaluations (outside of a functioning vehicle and outside of functioning network connectivity). As a result, the end state for a task is operationalized to be the completion of control inputs for the task sequence, or as indicated by the test subject, as appropriate to accurately measure the duration of the task. Example: A destination entry task ends with the final control input that initiates way-finding. This is an example of a task that ends with the final control input (p. 65).</p>	
Transitions Between Tasks	<p>One source of workload in a driver’s interactions with an advanced information system is making transitions between tasks in different parts of the system (e.g., moving from navigation functions to radio functions). As such, for</p>	<p>Example: At system start-up, the telematics display default screen shows the audio system (<i>the top-level screen for the audio system function</i>). When evaluating a</p>

	<p>purposes of evaluating compliance with the principles and criteria below, transitions between major system functions (e.g., power-up default screen, NAV, phone, Internet, radio, etc.) should be evaluated and, when evaluated, could be treated as separate “tasks.” This method for determining which transitions to evaluate should help identify transitions that have a high expectation, real-world likelihood of consumer use (p. 66).</p>	<p>NAV task such as destination entry, one must first evaluate the “transition task” of initiating NAV, starting at the audio system display; then one must evaluate the NAV task of destination entry starting with the first NAV display upon function initiation (p. 66).</p>
--	---	--

The definition of a “task” in the Alliance Guidelines resembles the one used in the Society of Automotive Engineers (SAE) Recommended Practice J2364 (2004), which served as an original source for the Alliance Driver-Focus Working Group. SAE J2364 defines a “task” as “Specific sequence of control operations leading to a goal at which the driver will normally persist until the goal is reached.” The standard provides the following example: “Obtaining guidance by entering a street address using the scrolling list method until route guidance is initiated.”

Chapter 2. Methodology

Overview of Interviews

Information was obtained from experts through an interviewing process. At a high level, this process consisted of the following six activities:

1. Candidate experts were identified and reviewed by NHTSA.
2. A final list of experts was prepared by the contractor and the list was approved by the NHTSA task order manager.
3. A script was prepared for the interviews and reviewed by NHTSA. This script is provided in Appendix B. Also provided are an e-mail invitation to interview (Appendix A) and an interview preparation packet for respondents' use before the interview (Appendix C). To assure protection of human research participants, the research protocol was submitted to and approved by the Virginia Tech (VT) Institutional Review Board (IRB).
4. After e-mail invitations to interview were issued and responses received, the over-the-phone interviews were scheduled, and the interview preparation material was e-mailed to each expert in advance of their interview.
5. A telephone interview was then conducted with each expert who agreed to participate. Two declined, producing a total of nine interviews. When experts granted consent, the interviews were recorded using a teleconference-based audio-recording technique (to permit accurate off-line analysis).
6. Key concepts and ideas were extracted from the interview transcripts and notes were coded in a data-scoring matrix for summary of main ideas across experts.

The step-by-step process flow for activities 3 through 6, above, is depicted in Figure 1. The details of methodology, analysis, and specific results from these interviews with experts are described in the remainder of this report along with the conclusions and recommendations that emerged.

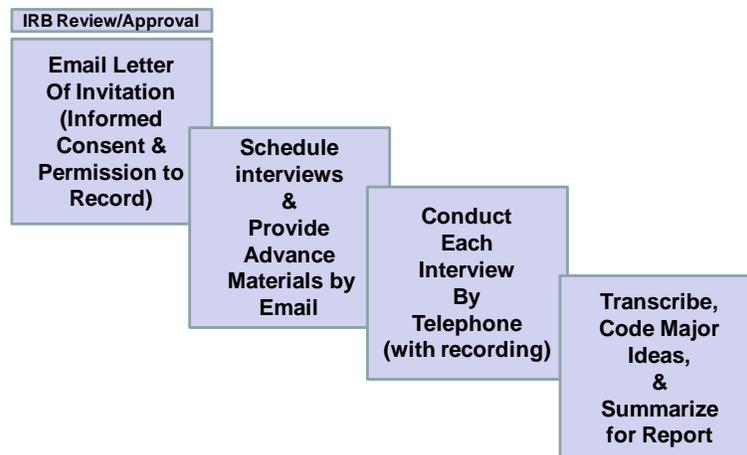


Figure 1. Overview of Interview Process

Sample of Experts Interviewed

Nine experts were interviewed. Of these, eight were *external* to the research team and its sponsors. The remaining expert was employed by the transportation institute carrying out this research but was *external* to the actual research team conducting the work described in this report. These experts were selected on the basis of their knowledge and experience in the area of driver distraction (and with the evaluation of devices for distraction potential). Three were drawn from major automotive companies, five from academic institutions, and one from an organization affiliated with government, regulation, or standards-making. Table 2 summarizes the types of experts who were included in the sample (protecting the exact identity of each, according to confidentiality guaranteed under IRB standards for research of this type).

Table 2. Sample of Interviewees

Type of Organization Represented	Country or Subtype of Organization Represented by Expert
Automotive Manufacturers	
	1 European Manufacturer
	2 U.S. Manufacturer
	3 U.S. Manufacturer
Academic Institutions	
	4 Transportation Research Institute
	5 Transportation Research Institute
	6 Transportation Research Institute
	7 Engineering Department
	8 Multi-Disciplinary Research Center
Government Regulation, Standards, and Research Organizations	
	9 Organization External to U.S.

Confidentiality and Compensation

Participation of experts in the hour-long interviews was voluntary and uncompensated. Informed consent procedures were followed (and review/approval through the VT IRB was obtained prior to initiation of the research). Confidentiality was protected through several methods in order to provide the experts with a comfortable and secure environment that would encourage an open exchange of information. The expert participants were identified by code only in data files, audio files, and on paperwork as a means of keeping personal identities of the interviewed experts confidential. Audio recordings and other records were maintained in secure storage at all times. A copy of the Invitation to Interview (and Request for Informed Consent) is attached in Appendix A.

Preparation of the Interview Script and Advance Packet

Because the individuals interviewed during this task were selected on the basis of their particular areas of expertise, the interview script was likewise tailored to elicit responses in focused areas of expertise. Questions and follow-up probes were prepared in advance. This guide was reviewed with the NHTSA task order manager and approved prior to the initiation of interviewing. A copy of the interview guide appears in Appendix B.

Administration of the Interviews

Phone interviews were conducted by a team of two project researchers. Within this team of interviewers, the same interviewer always administered the interview script (asked all the questions of the respondent) to assure consistency. The other member of the interview team took notes and was prepared to ask follow-up probes for clarification or further exploration of an answer. The interviews were recorded for subsequent analysis (when participants granted consent for such). When possible, full transcripts of the interviews were also obtained. In other cases (where recording was declined by a participant), the notes of the interviewing team served as the record of the respondent's comments.

Extraction of Ideas and Concepts From the Interviews

The intent of the interviews was to explore insights that the experts had into how “task” should be defined, as well as other topics explored in the interview script.

To maximize the amount of information and meaning that could be extracted from each interview, key units of meaning (consisting of main ideas and concepts) were extracted. These units were entered into a coding matrix to permit frequency counts to be obtained and to enable analysis of all the interviews so that common themes and unique pieces of information could be quickly identified. This process was accomplished in stages:

- Stage 1:** Compile notes, and obtain transcription of the interviews (those that were recorded).
- Stage 2:** Identify the key units of meaning. Annotate as appropriate with observations helpful to the creation of the scoring matrix.
- Stage 3:** Create the scoring matrix. The scoring matrix was structured by questions within the interview guide, with interviewees listed underneath and key units of meaning itemized by interviewee. Wording was retained from notes as exactly as possible.
- Stage 4:** Score each interview, entering the key units of meaning into the scoring matrix. Tallies across participants for certain concepts were then tabulated based on each question.

Recordings, transcriptions of the interviews, and use of two note-takers were used to control for issues that can arise in more traditional interviewing methods that can compromise the quality of data captured. Interviewing that depends exclusively on note-taking by a single interviewer can be heavily reliant on the note-taking ability and workload of the interviewer, so this method can miss relevant data. Further, it can be subject to the expectations of one particular investigator (of

which the interviewer may not even be aware) that may influence what is and is not captured in the notes.

Therefore, two methods were used to assure quality of data capture. First, by recording interviews when respondents consented, an accurate record was captured. Second, by having a note-taking member of the interview team (in addition to the interviewer) – an independent set of notes was created as a record of what was said, and this set could be used together with the interviewer’s notes to create a converging record of what was said in each interview. These measures provided a quality-control method to offer the best possible means of assuring that all key units of meaning had the opportunity to be identified and captured, not just those noted by a single investigator during an interview. This method helped to assure completeness and freedom from bias. Out of nine interviews, eight experts gave consent to be recorded. However, the conference-line recording mechanism failed to capture the interview for one of these eight, so full transcripts were obtained from seven of these eight interviews, and notes served as the data record for the remaining two.

Because the main purpose of the interviews was to help discover previously unreported or unpublished thoughts or ideas about how to define a task, the most appropriate data extraction technique for scoring the interviews was a form of “open coding.” This is a procedure in which units of meaning are identified without any restrictions or purpose other than to discover nuggets of meaning. It is applied in order to find the unexpected or new ideas in a set of transcripts. By applying open coding development in a stage that is *separate from* and *prior to* the actual scoring of the transcripts, it is possible to harness the “discovery” value of open coding while at the same time providing rigor in consistently coding all transcripts (in Stage 4, after the units of meaning have been extracted and established).

Analysis Procedures

In the past, it has sometimes been difficult to deal with qualitative data such as that which results from interviews, particularly when organizing and analyzing large quantities of “meaning units².” However, Qualitative Data Analysis allows the data to be systematically organized, and it permits exploratory data mining procedures (discussed below) to be applied, including the selection of particular subsets of data for analysis and reporting. This facilitates the finding of patterns in a set of qualitative data.

Because the data set was quite small, VTTI used a highly simplified form of qualitative data analysis, using simple frequency counts by units of meaning where possible to identify common themes in the interviews. Similarly, simple frequency counts were used for identifying new and unique information. *Higher frequencies* helped identify a *recurring mention* of ideas, causes, outcomes, etc.; *lower frequencies* helped identify *unique and new* pieces of information.

² A unit of meaning is a word (often called the node or keyword) plus all those words within its contextual context that are needed to disambiguate this word. An example would be: the word fire is ambiguous. Therefore it is not a unit of meaning. In connection with the noun enemy it becomes part of the collocation enemy fire, meaning “the shooting of projectiles from weapons by the enemy in an armed conflict”. This collocation is (under normal circumstances) a unit of meaning. “Traffic jam” would be another example of a unit of meaning. This information is based on the Web page <http://uk.answers.yahoo.com/question/index?qid=20090717073719AAaDgdV> on March 27, 2012.

Chapter 3. Results and Discussion for “Task” Definition

Across the interviews with experts key ideas and themes as well as unique thoughts (or singletons) were identified. These results are described below.

In considering these results, it is important to bear in mind that the scoring techniques extract units of what was *actually said* by a small set of experts (their actual spoken words). However, in spoken discourse, there is much meaning that lies unspoken and implicit. Also across the interviews, there are meanings that are not articulated but can be identified as *emergent from the set of interviews*. The reader should bear in mind that in the methods used here, only the spoken words of experts are captured by the scoring techniques. Therefore, in a few instances where it was appropriate for assuring that full meaning was conveyed (including unspoken meanings), a few inferences about implicit meanings have been included. These are explicitly marked with brackets and identified as *implications* where they were especially pertinent to expert responses to a topic (and could be identified with confidence from the set of interviews). In some instances, author notes are added and identified as well.

Another separate issue that requires mention is that sometimes a particular topic leads to a large number of “singleton” responses (and little or no convergence on key ideas or themes that could be readily summarized). When that happens, singleton responses can simply be listed and included for convenient review (since sometimes individual ideas have relevance or may prompt further ideas toward improved definitions or approaches). In other instances, a sample is so small that it constrains the degree to which convergence among experts has the opportunity to emerge. This makes the identification of truly common themes within the findings challenging and increases the number of singleton responses. As this sample was small, the reader is advised to bear these issues in mind.

Key Ideas and Themes

Several key ideas and themes emerged from the interviews. These are presented within the structure of the interview script, which was organized around four major topic areas:

1. Definitions of “task” currently in use;
2. Other definitions that are used in conjunction with the definition of “task” (e.g., “goal,” “subgoal,” “subtask”);
3. Alternate definitions for “task”; and
4. Task taxonomies.

Results from the first two areas of the interview are described in this chapter, with the results from each of the remaining areas addressed in following chapters.

Definitions for “Task” Currently in Use

Within the first topic area of the interview, several questions explored experts’ views about the current definitions of “task” (in the context of secondary activities while driving). Results are organized by themes.

Theme 1: The task definition in the Alliance Guidelines is satisfactory, usable, “good enough”; a “consensus” definition.

In the Alliance Guidelines, “task” is defined as “a sequence of control operations (i.e., a specific method) leading to a goal at which the driver will normally persist until the goal is reached.”

Experts responded to the questions (See Appendix B for complete list of questions):

- **How good is the Alliance definition? Is it the same as--or similar to—what you use in your organization?**
- **If not identical, what definition does your organization use?**

In response to the above question about how good the definition of “task” is that is used in the Alliance Guidelines, responses were as follows:

All (nine) experts indicated *in some form* that the current definition (in both the Alliance Guidelines and in SAE J2364):

- Is **usable; satisfactory; “good enough”** (six with explicit comments and three more by implication from the fact that it is similar to what is used in the experts’ organizations). The definition is not perfect, they said, but it is usable, it is not capricious, and its main points have been agreed upon through consensus within the discipline and industry.
- One expert pointed out that it is “. . . necessary to have a way to parse the functionality of the system in order to evaluate it. This task orientation is sensible.”
- The experts indicated that this definition has some important features, including:
 - The concept that a task is determined by a goal of the driver;
 - The concept of a sequence of operations (though some note that while “control operations” are the “core part” of a task, it may be appropriate to capture *other* human operations that are also integral to the task [e.g., glances] that may lead or follow control inputs);
 - The concept that humans tend to persist toward a goal until they reach it; and
 - The fact that the definition has some flexibility but not too much.
- All nine experts indicated that this definition is *basically* the definition they use for their work (though some indicated that they use an earlier starting point or a later end point):
 - It is not perfect but is usable; and
 - That its main points are agreed upon.

Theme 2: There are no better definitions of “task” readily available.

In discussing the current definition, experts said that:

- **There are no better definitions** that they could identify.
- No expert was aware of a better definition that has been written about or published.
- All indicated that the Alliance Guidelines definition had been “worked over pretty carefully” by “different experts” and represented a “common balanced view.”

- Experts expressed the view that there are well-thought reasons for the way in which this definition is worded. (See the list of positive comments that were provided in the left-hand column of Table 3).
- Two experts were aware that SAE J2364 was the original source of the definition; some thought that *other* guidelines “must” include other definitions of “task” (such as the European Commission’s Statement of Principles [e.g., 2007] or the Japanese Automobile Manufacturers Association Guidelines [2004]). However, these other guidelines were examined, and no other definitions of “task” were found in them. (See Chapter 4 for more information about this).
- One expert reported a preference for his own improved version of this task definition and mentioned that his version would be more compatible with the human-computer-interaction (HCI) literature. This mention of the HCI literature as having used definitions related to task in a more sophisticated (albeit different) way prompted the authors to examine the HCI literature for a review of what is provided there.
- Another expert mentioned that although adherence to the Alliance Guidelines is first priority, this expert’s organization also considers three additional items: (a) quality of the human-machine interface (HMI), (b) how frequently the task will be performed, and (c) how relevant a task is to driving.

In their responses about the adequacy of the current definitions for “task,” experts did generate positive and negative comments. These comments are shown in Table 3. These were singleton responses. The positive comments identify elements of the task definition that should perhaps be retained as an improved definition is developed by NHTSA. The negative comments may identify areas where enhancements could be explored. However, see also experts’ responses to a direct question about whether the definition needed improvement (see Theme 3).

Table 3. Positive and Negative Comments From Experts About Current Task Definition

Positive Comments	Negative Comments
Captures the notion of people persisting toward a goal	Has some vagueness regarding key concepts: <ul style="list-style-type: none"> • What does “normally persist” mean? • What is “normal”? What if someone interrupts a task to attend to something else? How is this treated during measurement?
Has some precision	Is too narrow (1 expert), and has some vagueness (1 expert) for: <ul style="list-style-type: none"> • Start and end points of a task • Should include activities/glances/time spent searching for, finding, and initiating the task, not just from first button press onward. The same is true at the end of the task (i.e., should include activities/glances/time driver spends confirming task is finished/completed, not just final button press). • Author’s Note: This comment applies also to the definitions of “start” and “end,” not solely to the definition of “task.”
Provides some freedom and flexibility for application to real systems but not too much. (Some degree of vagueness/flexibility is good and may be necessary since real systems have variability in their structure and operation, and the definition must be able to accommodate tasks in all of them.)	Has some vagueness and may need additional definitions for: <ul style="list-style-type: none"> • What a goal is? • What is meant by the term “control operations”? (What does it cover? Not cover?)
Can be used together with supporting terms/definitions	Not clear how a “sequence of control operations” is linked to a goal

The concept of a “testable task”

One expert explicitly suggested that there may be a need to identify the definition of “task” as the one for a “testable task” (since the definition’s purpose is to standardize the way tasks are identified and used for testing device adherence to guidelines about distraction). Several other experts alluded to this need as well, but did not state it explicitly. The expert who identified it directly in his interview was indicating that there may be a need to explicitly distinguish the definition for a “task” as it is used for *testing* distraction (relative to a set of guidelines or limits) from more general, common parlance about “tasks.” In fact, he suggested that what may be needed for guidelines/tests about limiting distraction is a definition for a “testable task” (rather than a task definition that will work for accommodating all task types). The notion of a “testable task” is that it would be one that is defined more narrowly in order to provide for consistency and repeatability across system architectures, task types, and organizations. It may thus encompass less breadth (may not be able to accommodate all naturally occurring tasks) but would have more precision, clarity, and repeatability across all testing.

Another issue that emerged from the interviews was a distinction between the needs of research and the needs of product evaluations. Two of the experts mentioned that *for research* there may be a desire to have a task definition that is broader and more encompassing of behaviors in order to capture the full spectrum of human processes that science is interested in explaining. In contrast, the needs *for product evaluation* may require specificity and a narrower focus to assure

that the “object of measurement” (the task design as enabled by a device) is truly something that can be affected by the design of the task/device (as opposed to other, broader factors of the driving environment outside the control of the manufacturer). Thus, these experts commented upon the fact that the definition of “task” for product evaluation (a “testable task”) may well need to be different than the definition of a “task” for research. They also indicated that these two venues (research versus product evaluation) may differ in other important ways (beyond the definition of task) since their objectives are different.

One expert’s variation on the definition for “task”

As mentioned previously, among the interviewees was one expert who volunteered his own preferred definition of a task. It was:

A task consists of a GOAL and a specific METHOD for achieving it.

According to this expert, such a definition “provides closer compatibility to the Human-Computer-Interaction literature.” The key point made in this expert’s discussion is that a task is tied to a method of reaching the goal (a specific way of accomplishing the task goal). If a different method is used to reach the goal, it constitutes a different task even though the goal may remain the same.

Information From the Task Analysis Literature

As mentioned previously, the reference to HCI literature by the aforementioned expert prompted the authors to examine the task analysis literature and the HCI literature for further insights into how to define a task. Key observations in connection with this are as follows:

- The task analysis literature has already served as an indirect source for the existing definitions/terminology.
- Hierarchical task analysis (HTA) was introduced in the late 1960s as a formal discipline (Shepard, 2001).
- The HCI literature built upon some of this terminology (the seminal work was that by Card, Moran, & Newell, 1983).
- Cognitive task analysis evolved around the year 2000 (through the application of cognitive architectures for modeling behaviors).

A review of the task analysis literature, starting with HTA, reveals that it has been widely used in human factors practice since it was first proposed in the 1960s (Shepard, 2001). It has led to the development of various techniques for conducting task analysis (Jonassen, Hannum, & Tessmer., 1989). Using these techniques, tasks can be analyzed along several dimensions. Common results from task analyses include the decomposition of tasks into subtasks and task elements. A task’s resource requirements can also be identified analytically, as can the simultaneous or successive nature of task components and their durations. In addition, these techniques are also sometimes used in conjunction with ability and skills analysis or assessment techniques (e.g., Fleischman, 1975, 1991).

The HCI literature began to emerge in the late 1980s with the Card, Moran, and Newell (1983) book, *The Psychology of Human-Computer Interaction*, being one of the first to widely introduce

the terminology. Chapter 9 of their book discusses the terminology for “overall tasks” and “unit tasks.” This literature laid the groundwork for the subsequent development of cognitive task analysis.

A useful example from cognitive task analysis, which reveals the way tasks are structured and defined within that approach, is depicted in Figure 2. It comes from a model called COGNET developed by Zachary, Ryder, and Hicinbothom (first published in a book, *Cognitive Task Analysis and Modeling of Decision Making in Complex Environments*, issued in 1998 then re-issued in 2000) for a domain other than driving. In COGNET, “. . . it is assumed that information processing activity is presumed to occur through the activation and execution of chunks of *procedural knowledge*, each of which represents an integration or compilation of multiple lower-level information processing operations around a domain-specific high-level goal. This combination of the **high level goal** and the **procedural knowledge needed to fulfill it** are referred to as a *cognitive task*” (p. 14, emphasis in bold was added by authors of this report). . . “The task is thus modeled as containing both the goal – and the knowledge-and-procedures for accomplishing it. *All the knowledge compiled into each task is activated* whenever the high-level goal defining that task is activated. Each task-level goal includes metacognitive knowledge that defines the contexts in which that task is relevant. This metacognitive knowledge is simply a description of the contexts under which the goal should be activated. Thus, the high-level goals are activated according to the current problem context, as defined by this metacognitive ‘trigger.’ In addition to this trigger, another piece of metacognitive knowledge defines the relative priority of that goal in the current context, so that attention can be allocated to the goal with the highest priority given the current context. This second piece of metacognitive knowledge is called the “priority expression” (p.14). These common features (described by the two quoted excerpts above) provide the structure for describing a cognitive task in COGNET.

TASK = High-Level Goal and Procedures/Procedural Knowledge to Fulfill it

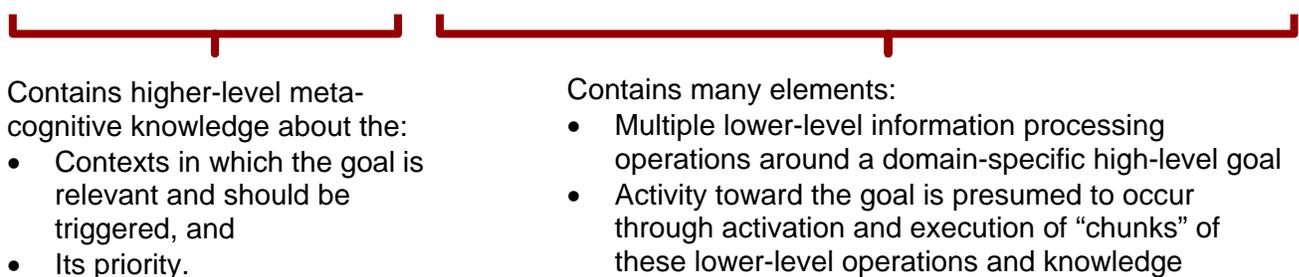


Figure 2. Structure of Typical “Task” Definition From Cognitive Task Analysis (Adapted From Zachary, Ryder, & Hicinbothom, 1998, 2000)

Discussion of Example From Cognitive Task Analysis

It is worthwhile to further discuss the internal structure given to a task within cognitive task analysis. This structure may be useful in considering how to describe a new, improved definition of a “testable task” for use with guidelines and testing for distraction limits for devices. The internal structure used in cognitive modeling (again, from Zachary, Ryder, & Hicinbothom,, 2000) can be described as follows:

“Each task has two main parts:

- The task definition, and
- The task body.

The task definition identifies the high-level goal involved and a specification of when that goal is activated and the priority it will have when activated.

A cognitive task is defined in the following form:

TASK <task-goal-name> ... *Activation Condition /Priority (formula)*. . <task body>

The task body is a hierarchy of lower-level information processing operators, based strongly on the GOMS (Goals-Operators-Methods-Selection Rules) notation of Card[, Moran & Newell] (1983), but with customizations to allow for:

- Manipulation of concepts on the blackboard (a formalism used in cognitive modeling);
- Evaluation of GOAL conditions on the basis of the blackboard *context*; and
- Interrupting and suspending the current task.”

One of the key differences between the early traditions of HTA and the later traditions of cognitive task analysis thus lies in the amount of “human interaction and human processes” included in the definition of task. In the early days of task analysis, tasks tended to be defined strictly in terms of elements of the task environment (in terms of what buttons, pathways, and operations had to be used on the machine interface to perform the task and reach the goal), and these were separated from the human operator (Shepard, 2001). The task thus reflected the design of the device and its interface. In more recent approaches (e.g., cognitive task analysis), there is a pronounced shift toward viewing the human and the machine as an interacting system in defining the task and to include not just the actions taken on the machine or device but also the resources and inputs required by the human to make those actions; to include, in other words, the *entire* human-machine *interaction* accomplished through the task. In this view, the human and the machine interface are considered as an *interacting* unit. This raises the issue of where to draw the line between machine and human in a definition of task (for purposes of testing a device-enabled task), and results from subsequent portions of the interview may bear on this.

Theme 3: The Alliance definition can be improved upon.

Experts responded to the probe questions:

- **Does this definition need improvement? If yes:**
 - **What leads you to conclude this?**

- **What improvements would you suggest?)**

Expert-Suggested Improvements to the “Task” Definition (Singletons, with some paraphrasing):

- Simplify/tighten the language.
- Separate the definition of “task” into “core definition” and supplementary explanation.
- Add something about a task involving a specific method to achieve a goal.
- Use more examples, photos for clarity.
- Integrate the concepts of frequency-of-task-use and relevance for the driving task.
- Add something about the (primary) driving task – about vigilance tasks – and some acknowledgment that tasks are often concurrent with parallel processing going on in driving.
- Address the issue of tasks that might be stopped and started again; concatenated tasks.

Expert-Suggested Improvements That Apply to Definitions Used With “Task”

- Modify start and end states. The notion here is that leading glances (or other behaviors) and trailing confirmatory glances should be included.
- Address the issue of concatenated tasks. The concerns here are twofold: (1) how interruptions that normally happen during tasks should be handled during testing, and (2) how to reasonably address the issue of tasks strung together, one after the other, or concatenated into a string.

Theme 4: Differences in interpretation/application may be smaller than thought.

At the beginning of this undertaking, there was interest in whether the current definition of “task” was robust enough that it is similarly interpreted and applied across the industry. The interview was structured to explore this question in two ways: (1) by asking experts directly about whether the definition is similarly interpreted and applied across the industry, and (2) by having experts demonstrate their interpretation and use of the definition on four tasks. This allowed a comparison of what experts “said” about using the definitions similarly and what experts “did” when they actually applied the definitions to real tasks. The results of these two approaches are described below.

What interviewees said

Experts responded to the probe questions:

- **Do all organizations like yours (who use this type of definition) interpret it like you do and apply it similarly?**
- **If not, what differences are you aware of in how tasks are defined between organizations?**

Regarding differences among organizations in the interpretation or application of the definition for “task,” the experts said:

- All had very little information about differences among organizations (that there is not an open exchange on this topic within the industry due to the proprietary nature of business practices among competitors).

- Five experts, however, conjectured that if there were differences among organizations they would likely be in the following areas:
 - Treatment of task versus subtasks (inappropriate division of whole tasks into smaller entities [e.g., subtasks] for testing).
 - Differences in treatment of “start state.”
 - Differences in treatment of “end state.”
 - Differences in product development processes (rather than definition use). This difference in process, it was explained, could lead to different product outcomes (which might seem as if they could be due to differences in definition use when that is really not the root cause). An example of this that was cited was: A manufacturer tests some tasks very early for visual demand (using the task definition correctly), and the tested tasks meet guidelines. However, the manufacturer does not re-test, verify, or validate the tasks on a later prototype or saleable vehicle later in development (but instead assumes that the design has remained unchanged). However, in reality, changes were made during development that took the device out of compliance with the limits on visual demand. Then, when it is tested by an outside group the product no longer meets visual demand limits. The reason is that manufacturer’s original process tested tasks only once and neglected to verify/validate them again later in the process, not that definitions were improperly used/applied.

How interviewees used the task definition

- Interviewees were asked to demonstrate how they used the set of Alliance definitions related to task by marking the start and end states of four tasks:
 - Two tasks were based on a task analytic breakdown (destination entry, manual cell-phone dial)
 - Two tasks were based on a behavioral record of driver responses (entertainment task, address book task)
 - Each exercise included several very real dilemmas about which the interviewee had to make decisions:
 - Whether to base task start on control inputs only or to include leading glances
 - Whether to base task end on control inputs only or to include trailing glances
 - How to deal with subtasks
 - What to do about the common use of experimenter instructions (often used operationally) in defining start/end states

Regardless of the format in which the tasks were presented, all experts were as inclusive (or more so), in their application of the task definition to these exercises than was required by the Alliance Guidelines. These results show convergence in the use of the “task” definition within a step or two at the beginning of the task and a step or two at the end of the task.

Task exercises are shown in Figure 3 to 6. The results for individual tasks are shown in Table 4.

Table 4. Results of Experts' Use of Task Definitions for Actual Tasks

Task Name	Guideline REQUIRED Start - End	Guideline ACCEPTABLE Start – End*	# of Experts Reporting START at This Point		# of Experts Reporting END at This Point	
			Step 0/1	Step 2	Step 8	Step 9 or >9
Navigation Destination Entry Part A			Step 0/1	Step 2	Step 8	Step 9 or >9
REQUIRED by Guidelines	Steps 2 - 8			2 Experts	5 Experts	
ACCEPTABLE Under Guidelines		Steps 0/1 – 9 or >	7 Experts			4 Experts
Manual Cell Phone Dial Part B			Step 0/1	Step 2	Step 13	Step > 13
REQUIRED by Guidelines	Steps 2 - 13			2 Experts	7 Experts	
ACCEPTABLE Under Guidelines		Steps 1 – 13 or >	7 Experts			2 Experts
Entertainment Task Part C			Step 0/1	Step 2	Step 19	Step 20-22+
REQUIRED by Guidelines	Steps 2-19			2 Experts	0 Experts	
ACCEPTABLE Under Guidelines		Steps 0/1-21/22	6** Experts			7 Experts
Address Book Task Part D			Step 0/1	Step 2	Step 20	Step 22- 23+
REQUIRED by Guidelines	Steps 2-22			2 Experts	2 Experts	
ACCEPTABLE Under Guidelines		Steps 0/1-23	6 ** Experts			7 Experts

*The Alliance allows for the use of verbal indicators during testing such as “please begin now” to start the task and the driver’s verbalization of “done” at the “end” of the task.

**One expert abstained, indicating that he/she always marks the task’s “start” by using the experimenter’s command, “please begin now.” Because this was not shown on the behavioral record, the expert could not specify the “start.” Similarly, the expert uses the driver’s verbalization of “done” at end of task but only in instances where the finish cannot be clearly observed.

PART A of Definition Application Exercise. Below you will find an analysis of a task, decomposed into a numbered sequence of steps. Using the set of definitions from the Alliance Guidelines, please determine where the “TASK START” and “TASK END” would be demarcated for purposes of testing performance on this task *in your organization*. Please indicate your answer by filling in the blanks below, which are labeled *Task Start* and *Task End*.

TASK A	
Goal:	Enter destination of 1040 South Elm Street, Greenlee, Indiana
Task:	<ol style="list-style-type: none"> 1. Look for and find the system function button (a soft button) which brings up “NAVIGATION” functions 2. Press that system function button to bring up the main “NAVIGATION” screen 3. Press the (soft) button which allows for “Address Entry” 4. Press (soft) button for entry by “Street and House Number” 5. At prompt, enter name of state and city <ol style="list-style-type: none"> a. Either use keypad to enter ‘state’ using letter-by- letter entry -- and then ‘city’ using letter-by- letter entry, OR b. Select state from list at top of screen by touching it, and then selecting city from the next list that appears by touching it. 6. At prompt, enter name of street <ol style="list-style-type: none"> a. Either use keypad to enter name of street letter-by- letter, OR b. Select street from list at top of screen by touching it 7. At prompt, enter address <ol style="list-style-type: none"> a. Either use keypad to enter numbers of address, OR b. Select address from list at top of screen by touching it 8. Push “GO” so the map will display and route instructions will begin. 9. Look at map and receive first instruction.
Device:	Embedded navigation system with large color screen near top of center stack

Figure 3. Task Analysis for Destination Entry with Multiple Subtasks

According to Alliance Guidelines, the task should be defined as the sequence of operations **beginning at Step 2 and ending after Step 8**. Alternatively, the Alliance Guidelines allow for the use of verbal indicators, “please begin now” and “done” for “start” and “end” states, which would mean the task begins at Step 1 and continues through Step 8 or until the participant says “done” (which could be at Step 8, 9, or beyond).

PART B of Definition Application Exercise. Below you will find an analysis of a task, decomposed into a numbered sequence of steps. Using the set of definitions from the Alliance Guidelines, please determine where the “TASK START” and “TASK END” would be demarcated for purposes of testing performance on this task *in your organization*. Please indicate your answer by filling in the blanks below, which are labeled *Task Start* and *Task End*.

TASK B
Goal: Place call to person at 751-9367
Task: <ol style="list-style-type: none"> 1. Move hand toward phone, using eyes to guide its position 2. Push “Phone” to select phone functions 3. Select “dialer” 4. Recall phone number 5. Move hand and eyes to phone keypad 6. Enter digit 7. Enter digit 8. Enter digit 9. Enter digit 10. Enter digit 11. Enter digit 12. Enter digit 13. Push « send »
Device: Hand-held phone docked in vehicle with ear-bud in place prior to task

Figure 4. Task Analysis for Manual Phone Dialing

According to the definition of “task” in the Alliance Guidelines, the task should start at **Step 2 and end at Step 13**. Alternatively, if using verbal indicators of “please begin now” and “done,” the task would start at Step 1 and end at Step 13.

PART C of Definition Application Exercise. This time, we would like you to do something similar to what you did in Parts A and B (marking the start and end of the task), but instead of using a task analysis to represent a task, we would like you to make your determinations based on a record of driver behavior in the task. So below you will find a time history of responses to a particular task, with each horizontal line in the record illustrating a different type of response (glances, button pushes, etc.) over time (from left to right).

We would *still* like you to use the set of definitions from the Alliance Guidelines, and we would still ask that you please determine where the “TASK START” and “TASK END” would be demarcated within this test of performance data *in your organization*.

Please indicate your answer by filling in the blanks below, which are labeled TASK START and TASK END. Use the “ruler” under the figure to help fill in the blanks.

Entertainment Task (in an embedded system)

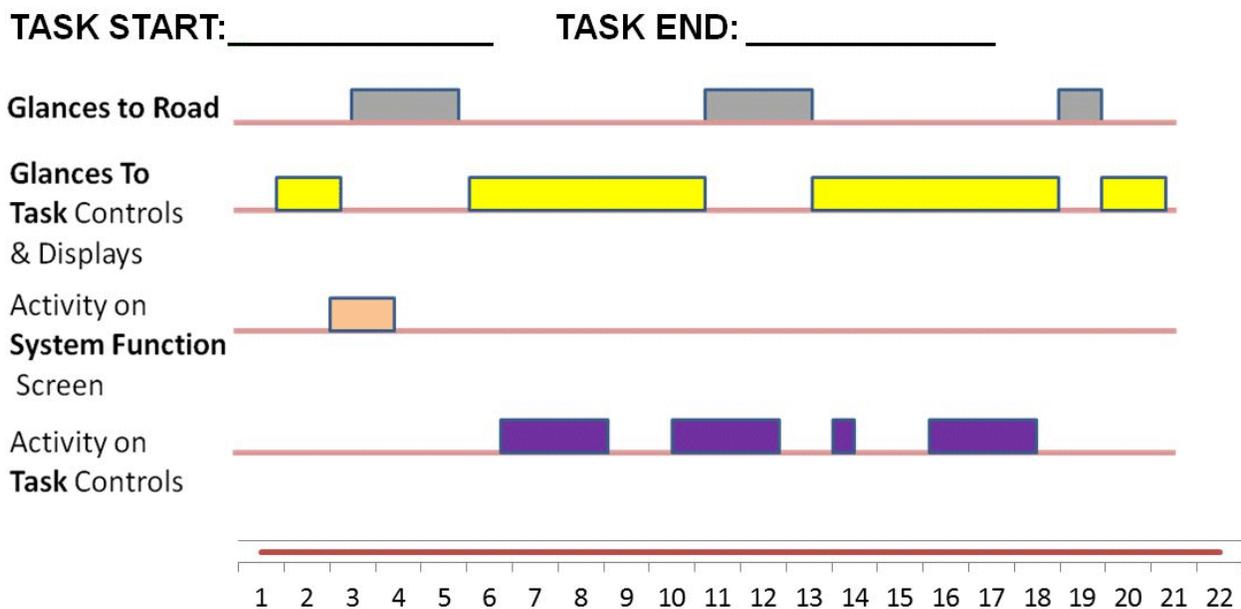


Figure 5. Behavioral Record for Entertainment Task

According to the definition of “task” in the Alliance Guidelines, the task control sequence **begins at Step 2 and ends at Step 19**. Alternatively, if using the verbal indicators of “please begin now” and “done,” the task would start at Step 1 and end at Step 21/22.

PART D of Definition Application Exercise. This time, we would like you to do something similar to what you did before (e.g., in Part C, marking the start and end of the task). We would like you to make your determinations based on a record of driver behavior in the task. So below you will find a time history of responses to a particular task, with each horizontal line in the

record illustrating a different type of response (glances, button pushes, etc.) over time (from left to right).

We would *still* like you to use the set of definitions from the Alliance Guidelines, and we would still ask that you please determine where the “TASK START” and “TASK END” would be demarcated within this test of performance data *in your organization*.

Please indicate your answer by filling in the blanks below, which are labeled TASK START and TASK END. Use the “ruler” under the figure to help fill in the blanks.

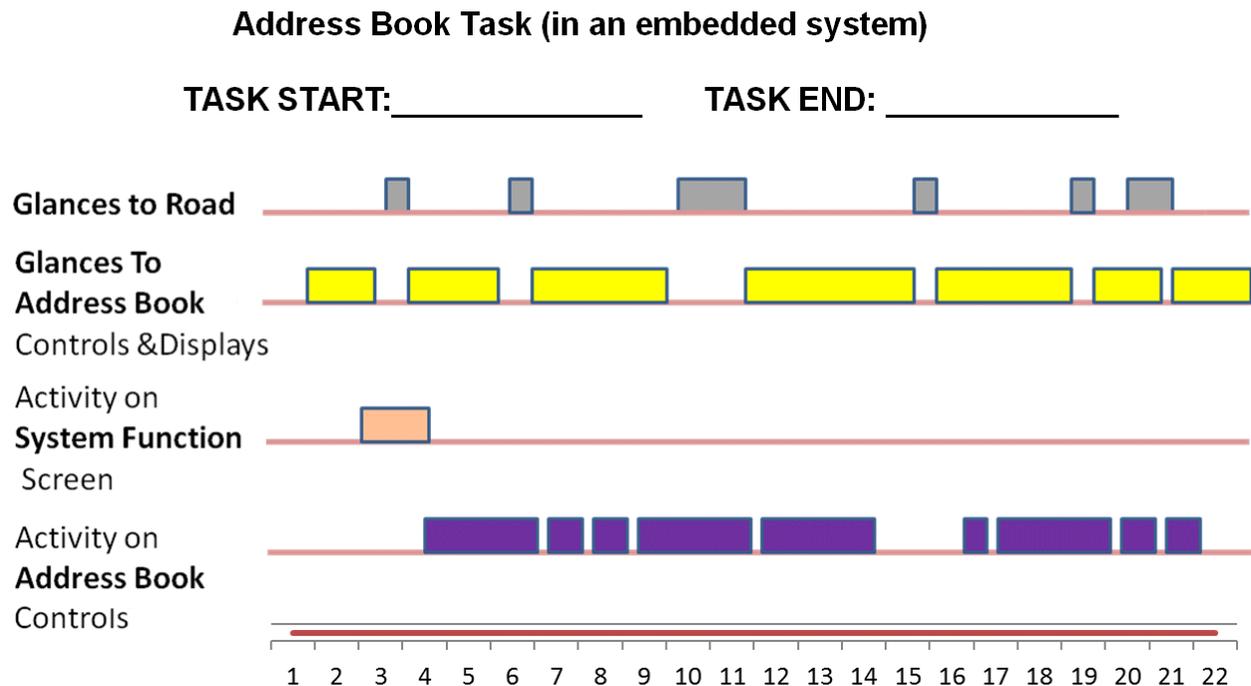


Figure 6. Behavioral Record for Address Book Task

According to the definition of “task” in the Alliance Guidelines, the task control sequence **begins at Step 2 and ends at Step 22**. Alternatively, if using the verbal indicators of “please begin now” and “done,” the task would start at Step 1 and end at Step 23 (the end point of segment 22, which is not labeled in the figure, but which was implied and used by participants).

Overall Summary of the Use of Definitions in the Exercises

As is apparent in Table 4, all experts defined the task epochs in a way that was as inclusive of human interactions as was required by the Alliance Guidelines definitions – or else was more inclusive than required (e.g., included additional leading or trailing eye glances in the task). In other words, the experts specified a start-to-end range for each task that was as broad as, or broader than, that defined by the Alliance document.

The responses of experts were not identical, however. There were some small differences among experts, including:

- Whether or not leading glances or leading hand movements toward task controls were included.
- Whether or not trailing glances were included.
- Use of verbal indicators (from experimenter and driver) in establishing “start” and “end” states of task.

The differences among experts were:

- One step at the beginning of a task and one step at the end.
- One glance at the beginning and one more at the end (up to a difference of two glances total).

These differences seem minor, but when accumulated over a task epoch, these differences could make the difference between a task’s passing or failing to meet a criterion, at least for tasks on or near a criterion’s borderline (e.g., on glance measurements).

Considering these findings:

- If a modified definition was to explicitly require inclusion of leading/trailing glances, it might be asked how important these seemingly small differences would be.
- Small differences of the type observed here would primarily affect tasks falling at the borderline of visual demand criteria.
- Some of the experts from organizations *other than manufacturers* believed (or expressed in the interview) that such borderline tasks tend to be rare and insignificant in the overall scheme.
- Automobile manufacturer experts considered that borderline tasks can occur more frequently than they would like during development, and that these borderline tasks can be agonizing in the overall scheme of evaluations conducted to limit distraction.
- At the same time, nearly all experts indicated that they are already including these “extra” steps/glances in their application of the definitions to test tasks, so a change of this type would appear (from their behavior in the interview) to be reasonably acceptable (if these experts’ behavior is representative of practice throughout the industry).
- One possible caveat is that experts may have been behaving under an effect for this interview where they exhibit behavior as it should be rather than as it is in actual practice in order to demonstrate their knowledge and expertise. This possibility should be considered. If such an effect were present, it would mean experts’ behaviors may be taken to indicate understanding but perhaps not actual application during testing and perhaps not readiness to accept changes in broadening the scope of the definition.

Three Key Issues That Arose in Discussion and Use of the Task Definition

Based on the comments of experts, there are three issues of sufficient importance to highlight for discussion. These are:

1. Purpose of a “test”:

- Experts reiterated that the needs of a product evaluation test differ from the needs of research. Although this was previously discussed in this report, this issue resurfaced when experts were explaining their applications of the task definition to the analysis of real tasks:
 - Research has as its purpose to “understand driver behavior” and may measure more broadly.
 - Product evaluation testing has as its purpose to verify adherence to limits/guidelines, and may measure more narrowly/precisely in order to perform assessments that are standardized across all organizations in the industry that must test in a similar manner.

- The two experts who mentioned this issue (that the needs of research and product evaluation are different) tended to agree that:
 - Research would generally need to be more inclusive of leading and trailing glances.
 - Whether or not to include leading and trailing glances (and reaching of the hand toward the device in advance of the first control input) in product evaluation was seen as a somewhat problematic issue.

- One reason this issue is a more difficult one in the product evaluation context was provided by one expert: It is necessary to test early and often during product development in order to achieve a device that adheres to guidelines. The testing tools used throughout the process need to yield answers that are consistent with one another. However, unfortunately, given today’s state of the art, the tools used early in the process have limitations that may mean that only a simpler (narrower) task definition can be used.
 - Specifically, techniques of today that are used early in product development (e.g., task analyses and predictive modeling tools) are typically centered on control inputs alone (and cannot yet comprehend glances or shifts of attention to the device).

 - Thus, if early predictive tools are used by manufacturers (and these early tools require that the definition of “task” be based on control inputs alone), then the answers from these tools will differ from answers obtained later in development (if the tools used later in the process are based on a more comprehensive task definition that includes glances, as well as control inputs). This would be a problem if early analysis provided different answers about the distraction potential of devices/tasks than did later testing because costs of changing device design late in the process are steep. This is particularly problematic if these changes are induced not because of any change in the device’s or task’s effect on distraction, but simply because the task definitions are not consistent across the types of tools applied for measurement in early versus later stages of product development.

Ideally, a task definition would be selected that is compatible with both those tool types used early in the process by manufacturers, as well as those used later in the process for

verification, so that answers are consistent throughout the development process. The only possible alternative would be to improve those tools used earlier in the process, so that they cover a broader task definition. It is not at all clear how these improvements would happen, but they would require time and resources, while product development would still have to continue.

2. Are the behaviors prior to the first button press and trailing the last button press clearly interpretable as attributable to the device/task interaction?

The questions are:

- Are these glance behaviors repeatable?
- Are they “real”? Do they happen outside of the test (in real driving)? For example, if a test participant glances at the device as soon as a tester begins to specify the task, is this glance behavior associated with the giving of the test instruction and thus an epiphenomenon of the test setting? Or do glances to the device also occur in natural driving as a driver prepares to undertake a task? One of the experts wondered if some empirical confirmation of these questions about leading/trailing glance behaviors is needed, and asked if there are existing data that already pertain to this.
- One study that has provided data relevant to these questions is a study of naturalistic use of infotainment systems reported by Angell, Perez, and Hankey (2008) and Perez, Angell, and Hankey (2010). In this study, drivers were given instrumented vehicles to drive as their own for a period of four weeks. Both video and driving performance data were acquired continuously. The study revealed that for all types of tasks in which drivers naturally engaged, they exhibited pre-task and post-task glances to the device with which they were interacting during the task. For example, Figure 17 shows the glance rate to the system of interest for several task types as a function of pre-task, task, and post-task periods (in the figure, the term “event” refers to the task). Figure 8 shows glance durations as a function of pre-task, task and post-task periods. Figure 9 shows the total eyes-off-road-time in seconds as a function of pre-task, task, and post-task periods. This research confirms that the leading and trailing glances to a task occur in natural settings and are not unique to, nor an epiphenomenon of, the test situation. These figures reflect actual driver behavior that occurs naturally and not just in the test setting.

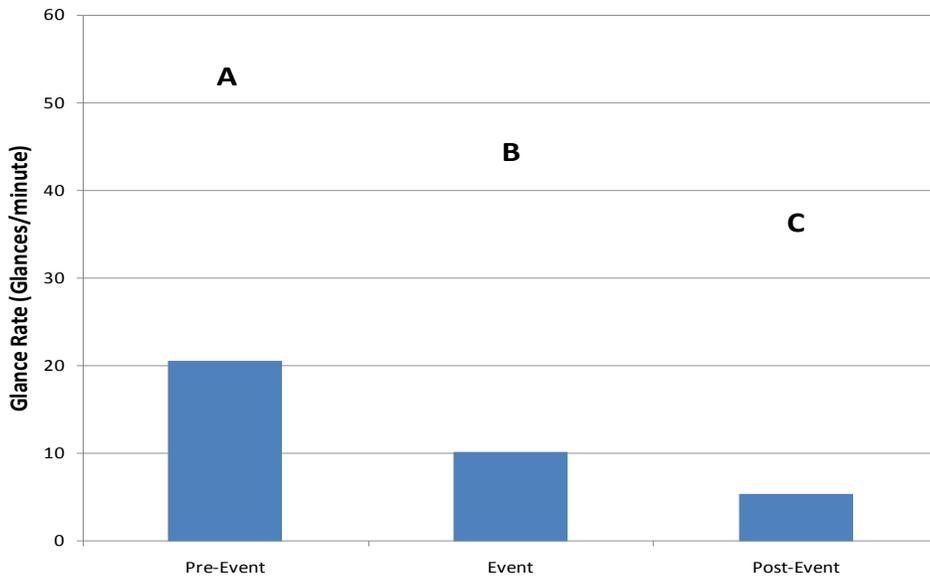


Figure 7. Glance Rate From Naturalistic Usage Study at VTTI During Pre-Task, Task (Event), and Post-Task Periods (Source: Angell, Perez & Hankey, 2008)

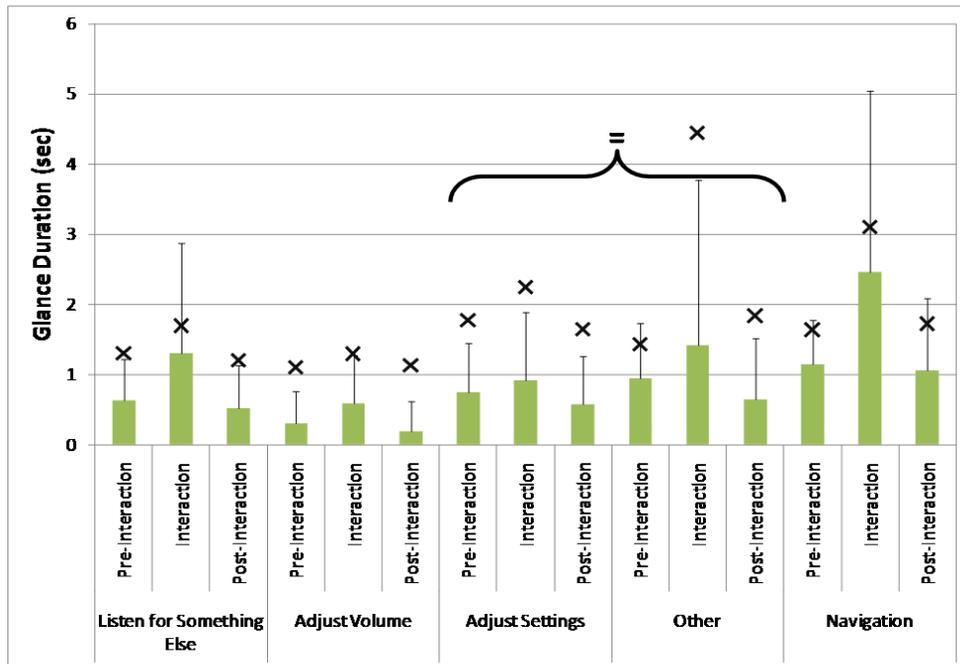


Figure 8. Glance Durations for the Pre-Task, Task, and Post-Task Periods of Several Task Types (Source: Angell, Perez & Hankey, 2008)

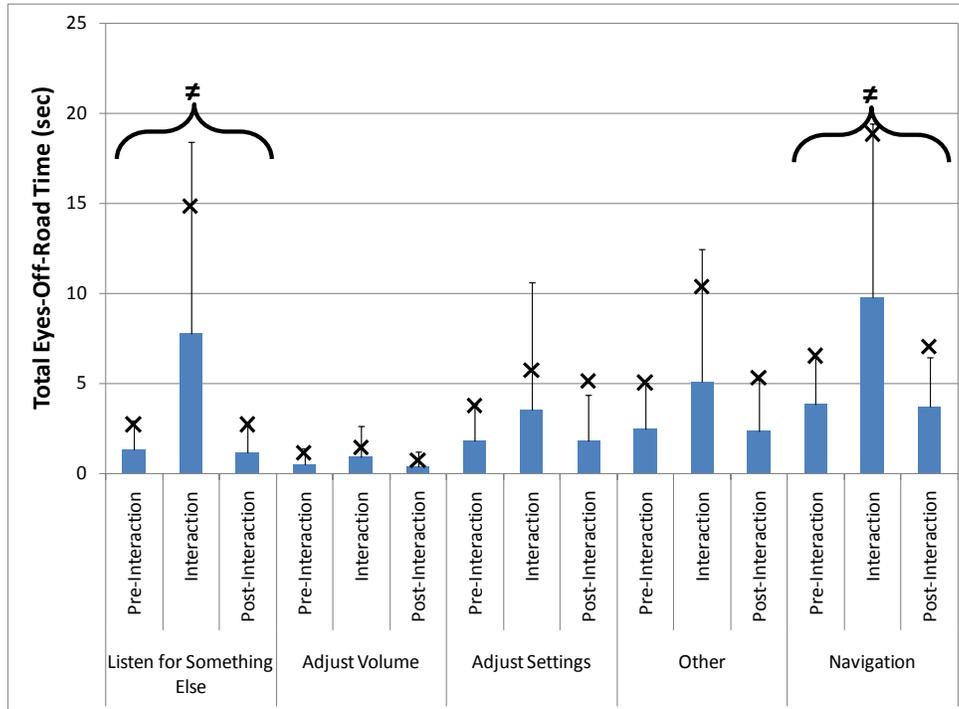


Figure 9. Total Eyes-Off-Road Time for the Pre-Task, Task, and Post-Task Periods of Several Task Types (Source: Angell, Perez & Hankey, 2008)

3. Should human behavior in the task definition be included? How much human behavior should be included? (That is, should the task be restricted to focus on the sequence of control operations that make up the core of a task or should it include glances or other behaviors that precede or follow it as well?) This issue was raised earlier in the report, and its importance emerges from the fact that many of the experts (though not all) tended to include both behavior and task elements in what they included for task measurement (e.g., in what they marked in their exercises defining start and end points of tasks). Considerations of how much human behavior to include in the task definition include the fact that:

- Early approaches to task definition separated the elements of the task environment (in terms of what machine buttons, pathways, and operations had to be undertaken by an operator to reach the goal) from the human operator (e.g., Shepard, 2001, on HTA).
- Later approaches to task definition included the human and the machine as an interacting system in defining the task (e.g., cognitive task analysis).
- The choice of which approach to use matters from two perspectives:
 - What the manufacturer can exert influence over (device design, operation). The manufacturer, of course, exerts influence most directly over the device and its design (its buttons, screens, labels, and other elements that may be used to perform a task). The manufacturer exerts only indirect influence over the driver's behavior (through the effects of the device's design in eliciting or attracting behaviors).

- How separable the device and user really are in terms of effects on driving performance and safety. It is not clear that the interaction of the human with the device can be separated, and treated as independent elements in their effects on driving performance and safety. The task seems to be an integral outcome of the human operating the device toward achieving a goal, and this human-in-the-loop system interaction is what generates effects that can have implications for safety.

Using these two perspectives, perhaps the most efficient way to optimize a device for a task is to consider both the elements of the task environment and the human operator as part of the task. This would suggest improving the definition of task to include them both (implying that both leading glances and hand movements toward the task, as well as trailing glances that are clearly attributable to completing the task would be included in the measurement that is defined for testing).

- Note, however, yet another concern in the choice of an approach for defining a task, as mentioned by one interviewee: the fact that not all organizations may possess equipment needed to acquire data about the human side of the interaction. In particular, the organizations may not all possess equipment or resources needed to acquire eye glance data with the efficiency to make its use practical.

Theme 5: Only one alternative definition rates similar to or better than the current one.

- Three alternative definitions for “task” were rated by the experts. These are shown in Table 5. They were rated by interviewees on a simple three-point scale of: “worse,” “same,” or “better” than the definition currently in use.
- **Alternative #1** explicitly allowed for “leading and trailing glances” and for the use of verbal indicators at the beginning and end of a task (e.g., “please begin now” and “done”). It was intended to be a modification of the existing definition that explicitly incorporated the practices of the majority of users.
- **Alternative #2** changed the definition entirely to an interruption-based notion.
- **Alternative #3** changed the definition to include “unbounded” natural searching and selection tasks.

Table 5. Alternative Task Definitions That Were Rated by Experts

Alternative Definition 1 for “Task”
<p><i>“A sequence of operations leading to a goal at which the driver will normally persist until the goal is reached . . . noting that these operations consist of any and all responses of the driver to the task (e.g., glances as well as control inputs, etc.). ‘Task start’ begins with the first activity on the task, and ‘task end’ finishes with the final response to the task. These initial and end responses may be glances rather than control inputs.” Note: To capture all responses, the “task start” during testing may be operationally defined as beginning as soon as the tester finishes saying, “Please begin now.” The final observable response to the task may be the driver saying, “I’m done” in a test.”</i></p>

Alternative Definition 2
<i>“A sequence of secondary task activities leading to a pause in the behavioral record during which the driver re-directs his/her attention back to the roadway and to driving for a period of time long enough to re-establish situation awareness (which typically takes 1 to 3 seconds, or perhaps more, depending on conditions).”</i>
Alternative Definition 3
<i>“A sequence of operations leading to a goal at which the driver will normally persist until the goal is reached . . . or until a driver need is satisfied.” (This is intended to accommodate “natural” tasks, in which drivers may persist in activity to satisfy a personal need . . . such as searching and searching for something good to listen to on a long drive, for example).”</i>

The rating results are shown in Table 6. As the table shows, while Alternative #1 was not overwhelmingly embraced, it was seen as the same or better than the existing definition by more than half of the experts (6.5; one expert was undecided between two ratings and was scored 0.5 for each). Perhaps the wording of Alternative #1 could have been improved since it should have been recognized as the same as the current definition with the modification that it includes leading and trailing glances, which the experts indicated they were already including in their application of the current definition. Thus, the experts should have recognized Alternative #1 as more or less the same as what is currently allowed and being practiced. However, it is interesting to note the high priority that these experts placed on careful wording and their skepticism about any change. Both Alternative Definitions #2 and #3 were rated as worse than the current definition by more than half of the experts.

Table 6. Number of Experts Rating Each Alternative Definition Worse, Same, Better

	Worse	Same	Better	Can't Respond
<u>Alternative #1:</u> Adding leading, trailing glances – allowing for use of verbal “start now” and “done”.	1.5	2.5	4	1 Needs to know exact measures of merit to give a rating on this one
<u>Alternative #2:</u> Complete change to an interruption-based notion	6	1	2	
<u>Alternative #3:</u> Adding search and selection to meet a driver's need (not just goal)	6	1	2	

Summary of Findings About the Use of “Task” Definition

The findings may be summarized as follows:

- Interviewees demonstrated a high degree of convergence on understanding, acceptance, and use of the current definition for “task,” which appears in the Alliance Guidelines.
- Experts also showed a high degree of convergence on the inclusion of leading/trailing glances even though the current definition does not require them to be included.
- However, there could have been an effect present in these results so some caution may be appropriate in interpreting and using these results.

Theme 6: Several sources of variation can give rise to differences in definition use.

Experts responded to the probe question:

- **When there are differences in definition interpretation/use, what gives rise to them?**

Experts offered the following (singleton) ideas about reasons for any differences that might exist in the way the definition of “task” is interpreted or applied. They indicated that differences would be due to:

- Different overarching purposes of evaluation (research versus product evaluation).
- Different overarching processes (some manufacturers test iteratively with verification of final hardware/software versus other manufacturers, which may test only once and early for some tasks).
- Different choices/philosophies regarding the choice of practices for Alliance Principle 2.1 A or 2.1 B verification of visual demand limits.

- Differences in equipment/tools available for testing.
- Different choices about how to select a representative task to test (e.g., length of address to enter when testing a destination entry task) and how to choose which tasks to test for a system.:
- Real differences in system architecture and function
 - Difficulties in mapping to “system function screen”.
 - Presence of response delays in systems.
- Limits of observation regarding “end state.” (It is not always possible to observe when the driver has finished the task; it may appear finished but the driver may not realize the end state has been reached.)

Theme 7: Actions can be taken to minimize some of these differences in definition use.

Experts responded to the probe question:

- **What could be done to minimize differences and establish common practice throughout the industry/discipline in applying definitions?**

Experts said:

- It may not be possible to achieve fully common use.
- Changes, even good ones, may come at a cost.
- Some ideas could facilitate more commonality. These include:
 - Provide more examples (e.g., from the top five types of systems).
 - Provide pictures/photographs to clarify the level of granularity at which the task definition should be applied.
 - Provide video.
 - Provide training? A possibility but perhaps it is a bit much?
 - Provide a system of checks and balances (to make sure companies are testing); ask for test results to be delivered.
 - Require documentation of what was tested and how, along with test outcomes.
 - [Authors’ Note: Not suggested by interviewed experts but an idea from the International Organization for Standardization (ISO) Working Group 8: Establish a calibration standard for test labs that will be testing device/task adherence to guidelines. One activity under ISO Working Group 8 is developing a standard practice through which test labs could become “calibrated” for testing by demonstrating that they are able to reproduce findings established by the standard. Perhaps a similar idea could be applied to NHTSA testing. Those companies performing tests could become calibrated for testing and demonstrate their ability to properly apply definitions and test procedures.]

Additional Ideas That Experts Suggested About Facilitating Common Practices

Experts provided additional comments regarding how to facilitate common practice:

- Resolve the underlying issues surrounding “interruptability” of tasks.
- Separate “transitions into tasks” from “tasks, themselves.” Experts suggested that this may help resolve differences in applying definitions. [Authors’ Note: It may resolve differences in definition application but it would also increase the testing

burden enormously. Under this suggestion, for any given task, there would be multiple different transitions into it that would need to be tested along with the task itself. It was this issue –testing multiple transitions into each and every task – that led the Alliance Working Group to identify the ‘system function screen’ as the starting point for tasks within a menu-based system. This solution means that at least one extra step, and perhaps more than one, is added to each task – as a representation of “average” transition difficulty within a system – and is used as an alternative to testing every possible transition into each task. (‘System function screen’ refers to the screen under which a set of related functions is grouped – such as ‘navigation,’ or ‘entertainment,’ or ‘communications’ – and from which a lower level task, such as “enter destination” would have to be entered.)

- Educate, explain the rationale, and demonstrate. There probably are or will be more non-human factors practitioners in the future as more applications and nomadic products are brought into the vehicle by drivers or integrated with the vehicle by manufacturers. These practitioners who are not formally trained in human factors may understand the definitions less well than their formally-trained counterparts, and may benefit from some supplemental education, rationale, and demonstration.

Ideas for *Mechanisms* to Achieve More Commonality

One expert spoke about mechanisms that might achieve more commonality, identifying three possibilities. Of these, the expert preferred the third alternative:

- Industry collaboration for purpose of a consensus standard. Cons: Too many philosophical differences. What should be executed? The U.S. model, German model, and Asian model do not reconcile into a standard that is uniform. Not sure how the idea of volunteer collaboration would work. Wish it would, but too idealistic. Also may not know enough to generate common circumstances and definitions.
- Regulation of industry. Distraction is not a pervasive driving condition like impairment. Regulation may promote very restrictive device/system design and functionality. Also may not address other alternatives such as portable devices or bringing in other devices that can be more dangerous than integrated devices.
- Use of a government-sponsored or independent third party to conduct device/task-testing (e.g., an extension of an Underwriter’s Laboratory (UL) listing-type test organization applying a New Car Assessment Program (NCAP) kind of evaluation). Third party would say, “This is how we process things and assess your system.” Beauty of this option is it is not government- or industry-driven. Has some accessibility and consistency. Could have some collaboration across university experts and endorse some kind of entity to fulfill this.

Concluding Comments

These comments concluded the results for the portion of the interview that focused on the definition of “task.” The next part of the interview explored definitions of *other* concepts and terms that are needed (along with the notion of “task”) when testing devices/tasks for distraction. The results from this part of the interview are described in the next chapter.

Chapter 4. Results and Discussion for Other Supporting Definitions Used in Conjunction With “Task”

Beyond the definition of “task,” the Alliance Guidelines set forth several other definitions for terms that are also commonly needed during testing of distraction. These terms include: system function, goal, subgoal, subtask, dependent task, start state, end state, and transition. In this chapter, the results from the interviews that pertained to these supporting definitions are described and discussed.

Theme 1: The supporting definitions are “okay” but not used as widely as expected.

Interviewees responded to each of the questions that are bolded below:

- **How good are these definitions as a set?** Results:
 - Eight experts said that they are “okay,” as good as anything they had seen; that the definitions have been worked over and work pretty well
 - One expert said that they could not comment and had little expertise about them
- **Is this set of definitions one that you use in your organization or is it similar to what your organization uses?** Results: Overall, this set of supplementary definitions was used less than expected. Specifically:
 - Two out of three experts from automobile manufacturers said “yes” they adhere closely to them; one said that they do not need or use these definitions.
 - One out of six experts from organizations other than manufacturers said “yes,” while three said that they do not need or use these definitions, and another two said “yes and no,” since they use a couple but not the rest.
 - On the whole, there was much less usage of these supporting definitions than of the “task” definition.
- **Are you aware of any other set of definitions that cover all (or nearly all) of these concepts or terms?** Results: Experts said that definitions in the two standards/guidelines documents should be considered. These were:
 - SAE J2364/J2365, and
 - European Statement of Principles.

The authors examined the two standards/guidelines documents to discover what supporting definitions are offered in each. Findings by the authors were as follows:

SAE J2364/J2365

The definitions provided in these SAE Recommended Practices are similar in many respects to the types and content of some of the definitions used by the Alliance. SAE J2364 was one of the original sources for the “task” definition in the Alliance Guidelines so it is not surprising that they would be similar. On the whole, the additional definitions found in SAE J2364/J2365 are compatible with the Alliance Guidelines’ supporting definitions but are not as fully specified.

Most salient among the terms that J2364 defines are the following:

- Goal (“system end state sought by a driver”).
- Task (“specific sequence of control operations leading to a goal or subgoal at which the driver will normally persist until the goal is reached”).
- Task Partitionability (“subdivision of a task into small time segments”).
- Computationally Interrupted Task (addresses system delays of 1.5 seconds or more).

Similarly, J2365 defines these terms:

- Goal (“system end state sought by a driver”).
- Subgoal (“change in system or device state necessary to achieve a goal”).
- Task (“sequence of control operations leading to a goal or subgoal at which the driver will normally persist until the goal is reached”).
- Method (“description of how a goal is accomplished”).
- Computationally Interrupted Task (task where there are system-induced delays of 1.5 seconds or more).
- Shortcut (alternative method by which a task may be completed more quickly).

European Commission Statement of Principles (ESoP)

This document provides a very different treatment of task and task-related definitions from that provided in the Alliance Guidelines:

- The ESoP distinguishes between the task of driving and the task of interacting with system displays and controls. ESoP calls for the allocation of driver attention to the task of interacting with system displays and controls to be compatible with the attentional demand of the driving situation.
- It goes on to say that the concept of “task” is controversial since the same task can vary substantially in terms of its parameters (e.g., in duration).
- Moreover, the ESoP says that a suitable definition of task is not available.
- It says that depending on the motivation and state of the driver, an interface with displays and controls may have a different effect; this is due to the fact that less workload is not necessarily better.
- The relationship among the components of the interface (complexity, intensity, duration, etc.), workload, and driving performance is not well enough understood.
- Systems that are designed in accordance with the ESoP should be such that the attentional demand of the system can be modified by the driver by choosing to interact (or not) and by choosing when and how to interact. This also means that the driver can anticipate the demand of the interface with the system.
- Thus, contrary to the experts’ expectations during this study, the authors found that there is both a different approach taken in the ESoP and that the ESoP provides no definitions of supporting terms related to the concept of “task.”

Theme 2: The set of supporting definitions can be improved.

Experts responded to these probe questions:

- **Are there any improvements to the definitions in this set that are needed?**
- **Are all the definitions clear?**
- **Are any of them confusing or ambiguous?**
- **What improvements would you suggest?**
- **What leads you to suggest each of these?**

Relative to the set of supporting definitions, experts suggested the following:

- **Delete this set altogether**, unless some organizations/companies find it useful. Use of these terms/concepts is not necessary. (One expert), or
- **Leave the definitions in this set as they are.** (One expert), or
- **Improve this set.** (Six experts; one additional expert abstained due to lack of familiarity). The following suggestions were made for improvement:
 - There are ambiguities in some definitions, but it may be that those ambiguities cannot be removed without a cost because sometimes you need some flexibility in order to apply these to a particular case and have it still make sense. (One expert)
 - Provide more complete linkage between terms (one expert):
 - Specifically a linkage of goal and subgoal with task and of subtask and subgoal with task.
 - Improve clarity (two experts):
 - Simplify/streamline wording, and
 - Provide examples (to make these supporting definitions more concrete and clear).
 - Add definition for self-pacing versus system/forced-pacing of tasks (two experts).
 - Add definition for system-induced delays in task (one expert).
 - If definitions change, consider that not everyone has the same equipment (one expert; e.g., not everyone has an eye-tracker or resources for reducing video so they cannot necessarily use glances as the start/end of a task).

Special Authors' Note [abstracted from interviewer comments]

Difficulties can arise for even the most conscientious application of these supporting definitions. These difficulties stem from the fact that:

- System architectures for information systems can vary a great deal (mentioned by three experts). This becomes apparent in trying to apply the definitions for the system function screen and when identifying “Start State” of the task for testing. Experts from automotive manufacturers already reported spending much internal dialog on this issue in order to reliably and fairly identify the system function screen from which a task should start (in order to provide consistent testing). This is not a problem with the definitions per se; it is simply an issue with the natural variation among the systems themselves.
- System complexity is expected to increase in the future. As it does, this issue is likely to become more difficult. Experts mentioned:

- Nomadic device integration.
- New technology features/functions.
- Intelligent/adaptive interfaces.
- Reconfiguring/personalized interfaces.

Results Related to Specific Terms Among the Supporting Definitions

A number of findings pertained to specific terms within the set of supporting definitions. These are described below.

Dependent Task Versus Subtask

The Alliance Guidelines define a “dependent task” as:

“There is a class of tasks (called “dependent tasks”) which may be distinguished from subtasks yet cannot be initiated until another task is first completed. Their “start state” is thus dependent upon the “end state” of another (antecedent) task. However, such tasks are to be treated as tasks (rather than as subtasks) for purposes of evaluating compliance of tasks with the principles and criteria below. They can be distinguished from subtasks by examining their end state (or goal state), which will usually be a driver-sought, system-independent state.”

Experts were asked to consider the definition for “dependent task.” They were asked:

- **Focusing specifically on the definition for “dependent task” (as opposed to “subtask”), is this definition clear enough?**
- **Can it be applied in a consistent way across organizations?**
- **Do you think it is?**

The results were:

- Two out of three automotive manufacturer experts thought the definition of “dependent task” was clear or as good as it could be (one automotive manufacturer expert thought it was not useful).
- One out of six non-manufacturer experts thought the definition of “dependent task” was good/clear; five thought it was unclear, confusing, or else disagreed with it:
 - Except when it was applied to route following.
 - The issue of multiple paths to a task was mentioned, and the associated issue that a task may not always stand in a dependent relation to another (it will depend upon the path through which it is entered).
 - [Author’s Note: This last concern is a non-issue since multiple paths to a dependent task underscore the separate nature of that task. Since dependent tasks are to be treated as separate tasks the fact that there are other paths besides the path of dependency from which to access the task confirms its ability to “stand on its own.” This contrasts with subtasks, which cannot “stand on their own”.]

Start State

The Alliance Guidelines define the “start state of a task” as:

“The start of a task under test is the default start screen for the system function under which the task falls. Every task within a system function must share the same start state for purposes of evaluation for compliance with these principles and criteria. An exception is made for tasks that can be initiated only following the completion of a previous task. For these dependent tasks, the start screen would be the end of the previous task.”

Experts were asked to consider the definition for “start state,” with the following questions:

- **Focusing specifically on the definition for “start state,” is this definition clear?**
- **Can it be applied in a consistent way across organizations?**
- **Do you think it is?**

The results were that:

- Two out of three experts from automotive manufacturers thought the “start state” definition was clear; the other thought it was not relevant for application of Alliance Principle 2.1b.
- None of the experts from non-manufacturers thought the “start state” definition was clear enough to lead to consistent application. They thought there would be issues with applying it consistently (e.g., mapping system architectures to the system function screen).
- Four of the non-manufacturers took issue with the “start state” definition. Comments included:
 - That it was artificial to have tasks start at system function screen, and some questioned the value of this.
 - That there are multiple points from which to start a task.
 - The possibility that adaptive interfaces may anticipate driver intent and configure the starting point of a task for the driver.
 - That perhaps transitions should be separated from tasks and tested separately.
 - The notion of incorporating leading glances and/or other leading behaviors prior to the first control operation (e.g., leading hand movements) continued to be mentioned.

End State

The Alliance Guidelines define the “start state of a task” as:

“For the purpose of testing to the criteria contained in this section, the end state of a task is the completion of the final manual input to achieve the driver’s goal, or as indicated by the test subject, as appropriate to accurately measure the duration of the task. This operational definition of task end state is necessary due to the fact that test systems may need to be used for evaluations (outside of a functioning vehicle and outside of functioning network connectivity). As a result, the end state for a task is operationalized to be the completion of control inputs for the task sequence, or as indicated by the test subject, as appropriate to accurately measure the duration of the task.”

Although experts were not specifically probed about the “end state” definition, some of them did comment on it:

- Two to three experts commented that it is not always possible to readily “see” the point at which a driver “finishes” a task.
- One felt that accepting driver verbal assertion that the task is completed is “not objective enough” and may be unduly influenced by the driver’s wish to perform well.
- However, two to three experts voiced the need for “confirming evidence” beyond the behavioral record and task analysis to identify when a driver has completed the task. They felt that the driver’s opinion of whether he or she had reached the end state of the task is one piece of confirming evidence that can be necessary and helpful in this regard.
- All experts want to make sure that properly identifying the point at which the driver “finishes” a task is performed in a careful and sensible way. The experts voiced the possibility that system-specific issues may require judgments to be made “in-context” with flexibility.

The Reasoning behind the Use of System Function Screen

As noted above under the discussion of the “start state” results, the non-manufacturer experts asked about the use of the system function screen in the “start state” definition, questioned whether it could be applied consistently, or questioned why it was included at all. An authors’ explanation of the reasons the system function screen was included in the Alliance Guidelines’ start state definition is provided:

- It was deemed necessary by the Alliance Driver-Focus Working Group to in some way reflect the driver’s need to “search and find” the start of a task within a system. However, there was recognition that there were infinitely many transitions into any given task. Starting at the system function screen was used as a way to realistically represent some “transition into a task” while reasonably bounding the infinite set of possible transition paths into any given task (according to the Alliance transitions *between system functions* are supposed to be separately evaluated as distinct tasks).
- Starting tasks at the system screen for their function type and evaluating transitions between system functions as distinct tasks are stipulated in the Alliance Guidelines as part of a comprehensive evaluation and ensure that all design-intent paths available to achieve the goal meet Alliance criteria.

These findings conclude the portion of the interview study regarding “supporting definitions” used in conjunction with the “task” definition during device testing.

In the next chapter, findings pertaining to “task taxonomies” are discussed.

Chapter 5. Results and Discussion About Task Taxonomies

Taxonomies of tasks are structures for organizing and classifying types of tasks based on similarities among them, as well as for organizing data about them (both test data and perhaps empirical data from the scientific literature). The last part of the interview with experts explored their responses to the use of different types of taxonomies as a part of the process for testing distraction, building knowledge about tasks of different types over time, and learning from the knowledge in such a way that the taxonomy might help guide certain decisions (e.g., about which tasks need to be tested or which ones could be judged based on past performance).

High-Level Findings About Taxonomies

Overall, the most important finding to emerge from the interviews about taxonomies was that few experts' organizations are currently using them. This breaks down as follows:

- Automotive Manufacturers:
 - Only one expert from an automotive manufacturer reported actively and routinely using a taxonomy and the taxonomy being used is being coupled with task analysis and modeling.
 - One other expert from an automotive manufacturer reported using some type of analytic technique, but it is not taxonomy-based.
 - Both of these experts stressed the need for, and importance of, a taxonomy for guiding what is “in scope” for testing.
- Non-Automotive Manufacturers:
 - None of the experts from non-manufacturers reported using taxonomies for the purposes discussed, though they could imagine doing so.

This lack of actual experience within the interview sample with use of task taxonomies for organizing distraction data or applying data on task types to distraction task evaluation limited the value of the interview data about taxonomies. Therefore, while the content of expert comments is reported here, the reader should treat these findings with caution.

Theme for Task Taxonomies: USE THEM ONLY WITH CAUTION.

In spite of the lack of use of task taxonomies, experts were familiar with taxonomies and understood for what they could be used. The overarching theme that emerged from their comments was that taxonomies should be used only with caution. Specifically:

- Many experts were concerned about:
 - Who would use the taxonomy; and
 - For what the taxonomy would be used.
- One expert expressed concern that a taxonomy approach does not fit well with Alliance Principle 2.1B comparisons (comparing a new task to reference task performance).
- Of six experts who do use a taxonomy now or who could imagine using one, their opinions about what it would be useful for were:

- Organizing data previously collected for tasks?
 - One Yes (and five speculating Yes)
 - Performing engineering analysis of new tasks early in design?
 - One Yes (and five speculating Yes)
 - Determining which new tasks may need to be empirically tested?
 - One Yes (and three speculating Yes)
 - Grouping tasks and selecting a representative task from the group to be tested (to streamline testing)?
 - One Yes (and five speculating Yes)
 - Applying redlines (or criteria) to tasks that are similar to ones in the taxonomy that have been previously tested?
 - One Yes (and three speculating Yes)
- Experts cautioned against:
 - Applying redlines based solely on a taxonomy unless it is applied to tasks that are well tested and technologies that are well vetted.
 - Applying predictive capabilities for new interface techniques and features based on prior expectations from a taxonomy. This is because past findings have not yet proven very repeatable and valid for new, novel interfaces. New, novel interfaces require follow-up verification testing with production-intent hardware/software. There is almost always a need for direct testing when something is new and novel.

Some additional findings about taxonomies included:

- One expert said he or she might recommend obtaining user feedback about a taxonomy before using it to determine if it matches a user's mental models of the system.
- A taxonomy based on the number of task steps might be very useful for human factors practitioners and developers though it may not be appropriate for regulators to use (the metric may not be robust enough for regulatory purposes).
- It will be very important to have some type of high-level taxonomy to define what is "within-scope" of the guidelines . (This was emphasized repeatedly by two experts.)

Ratings for Two Taxonomy Types

Experts were asked to rate two types of taxonomies (see Appendix C) on a scale that ranged from -4 to 0 to +4, where -4 was associated with the anchor ("Extremely inappropriate for use, a source of concern, a disbenefit"), 0 was associated with the anchor ("No added value"), and +4 was associated with the anchor ("Extremely useful, beneficial").

The outcome of the ratings was that neither taxonomy rated very highly.

- Taxonomy #1 (which was based on task structure and attributes) received an overall mean rating of +1.63, indicating that it had some use or some benefit.
- Taxonomy #2 (which was based on a blend of task properties and naturalistic data, including crash risk) received an overall rating of +1.38, which was lower than the first taxonomy but still judged to be of some value and benefit.

This finding, coupled with the fact that only one of the organizations represented by experts is currently using a taxonomy, suggested that perhaps the concept of applying a taxonomy to distraction testing would benefit from further development.

Separate Issue: Which Tasks Must Be Tested

[Authors' Note: In the absence of a taxonomy the issue arises of how a company determines which tasks should be tested. This is not an issue that was mentioned by the experts nor is it one that the experts were asked about directly. It is noted here simply as a consideration in the larger picture. Companies currently determine which tasks within a system must be tested. They may use engineering judgment, task analysis, a taxonomy, modeling, or some other approach.]

In new information systems, there are often hundreds of tasks that can be performed by a driver. Usually all such tasks would be expected to adhere to guidelines limiting distraction. Yet it is usually not possible to empirically test all such tasks on a routine basis (within the typically constrained resources available). Therefore, it would be helpful to have a fair and objective method for identifying which tasks to test (i.e., to identify which ones are likely to clearly adhere to guidelines and which ones need to be empirically tested to determine adherence). Comments from the interviews indicate that selection of tasks for testing varies among automotive manufacturers. This is a topic worthy of further consideration.

Other Issues Mentioned at Close of Interview

At the close of each interview, experts were asked if they would like to provide comments about any issues that not directly addressed during the interview. The topics (each a singleton) that they identified were:

- Research connections have not yet been made between measures of following headway and crash risk. This relationship needs research and precise measurement and quantification.
- As we move into the “apps era” details of these guidelines may be less understood by suppliers and non-human factors people involved in design and testing. As more interfaces are being designed by these groups, it is important to be aware of this.
- One additional important issue is the degree to which elements of a task need to be performed as an uninterrupted sequence.
- Long glances are an issue. Task definitions need to allow for a measure of these.
- A distinction between self-paced versus force-paced tasks needs to be added to the definition.
- Will cognitive load need to be assessed in addition to visual and manual load for visual-manual tasks?
- It would be desirable for the decision criteria within the Alliance Guidelines to be reviewed and addressed by NHTSA as it develops its guidelines. Of particular concern are those criteria having to do with decisions about glance duration and total glance times and the use of the 85th percentile in decision criteria.

These findings conclude those in the section of the interview dealing with task taxonomies and the wrap-up of the interview. The next chapter offers a summary of all findings and suggests preliminary recommendations.

Chapter 6. Summary, Conclusions, and Recommendations

Nine experts were interviewed about the definition of “task” currently used in the Alliance Guidelines and about definitions for supporting terms used in conjunction with the definition for “task” when devices are tested for adherence to limits on distraction. The objective of these interviews was to explore four major topic areas:

1. Definitions of “task” currently in use,
2. Other definitions that are used in conjunction with the definition of “task” (e.g., “goal,” “subgoal,” “subtask”),
3. Alternate definitions for “task,” and
4. Task taxonomies.

SUMMARY OF FINDINGS AND THEIR IMPLICATIONS

The results of qualitative data scoring for key themes and main ideas yielded salient findings in several areas, which may be summarized (along with the implications from each) as follows:

- **Current Task Definition.** In the Alliance Guidelines, “task” is defined as “a sequence of control operations (i.e., a specific method) leading to a goal at which the driver will normally persist until the goal is reached.” This definition of “task” seems widely accepted. The definition gives rise to fairly common practice (as evidenced both by what experts said and by what they did on a set of exercises in which they applied the definition to identifying the “start” and “end” points of tasks). One caveat that should be considered is that the exercises may have elicited an effect in which the experts strove to demonstrate their knowledge as opposed to what their organizations actually practice.
- **Alternative Definitions for “Task.”** There are not many already developed alternative definitions for “task.”
- **Possible Improvements to the Current Task Definition.** There are areas to consider for improving the current definition of task. They are: (a) whether the task should include leading glances or leading hand movements that occur prior to the first control operation in the task, (b) whether to include trailing glances (which occur after the last button press and may be associated with confirming that the task is finished), and (c) whether to explicitly incorporate (rather than just allow) the use of verbal indicators that identify task “start” and “end” states (e.g., the experimenter’s request to “please begin now” and the test participant’s response of “done” at completion).
- **Terms Which Support (are used in conjunction with) the Task Definition.** The interviews revealed that the set of supporting definitions of system-related terms (e.g., goal, subgoal, subtask, dependent task, transitions, and system function) is:
 - Not as widely understood (the experts working outside of the automotive companies were much less familiar with these definitions in the Alliance document than were the automotive experts).

- Not as widely used as the task definition (perhaps reflecting disagreements or a different range of practices or needs across the sample, which included both researchers and those at automotive companies involved in product testing).
- **Taxonomies for Tasks/Test Results.** The small study showed that currently:
 - Task taxonomies are not in wide use as part of distraction-testing programs (only one automotive company reported using a taxonomy as a routine part of its practices).
 - Two types of hypothetical task taxonomies shown to experts were not rated highly useful.
 - It may be too soon to propose a serious application of task taxonomies within the domain of distraction testing. Task taxonomies may require further development in order to be accepted by the community of intended users.
 - However, a task taxonomy could provide one basis on which to identify those tasks within a new information system that should be empirically tested for adherence to guidelines (versus those that are likely to already meet guidelines by virtue of their similarity to already tested tasks). Therefore, if no taxonomy is used, the issue arises of how a determination is made of which tasks within a system must be empirically tested – and it remains open for automotive companies. Each automotive company would need to determine which tasks to test in their own way as they are doing now (e.g., through task analysis, modeling, or engineering judgment). The issue of whether common practice among automotive companies is expected in this area or whether it is simply adherence of all tasks to the guidelines that will be expected should be considered in light of the findings about taxonomies.

Preliminary Recommendations

The findings from the interviews with a sample of experts led to the following preliminary recommendations:

- Adopt the term “testable task” instead of “task” in the NHTSA Driver Distraction Guidelines. Using testable task will emphasize that the NHTSA Guidelines are focused on tasks that can be tested. It will alleviate the need to encompass all possible secondary tasks that may be performed during driving. Focusing on testing will also facilitate providing greater precision in the definition of a testable task.
- Consider structuring the definition of “task” into two parts: a “core definition” (corresponding to the definitions contained in the Alliance Guidelines) and “supplemental information” (such as additional explanatory information and examples of testable tasks).
- Retain the current definition of “task” that is used in the Alliance Guidelines or slightly adapt it (following the suggestion of one interviewee):
 - A task consists of a GOAL and a specific METHOD for achieving it.
- To provide greater clarity, enhance explanatory and supplemental materials (such as specific examples of testable tasks) that accompany the definition of a task.
- Retain (from the current supporting set of definitions from the Alliance document) some selected key definitions (those for “start” and “end” state are particularly

important) but improve them in areas noted in this report, using examples to make the concepts concrete. This material could be made part of explanatory or supplemental material that accompanies the core definition of a testable task.

- Consider implementing a check-and-balance system to cultivate the commonality with which the definition of “task” is applied.
- Address explicitly the “scope” question (what types of devices and tasks are in scope) for guidelines and testing.
- Delay the notion of using a taxonomy for now, instead taking the time to develop the concept of a task taxonomy more fully (so it can be re-introduced later).
- Address explicitly what the expectations are about selecting tasks for guideline testing). Do all companies need to use the same method? Or do they simply need to make sure that all tasks that are within scope adhere to the guidelines?
- Address other issues that are “elephants in the room,” e.g., how to evaluate:
 - Tasks that are subject to interruptions, and
 - Task concatenations (the stringing together of tasks by a driver during driving, one-after-another).

These sessions with experts yielded key themes and main ideas that were useful in assessing the definition of “task” currently used in the Alliance document, exploring alternatives, and suggesting avenues for possible improvement. These key themes and main ideas may provide assistance as NHTSA moves forward to develop its guidelines and as it chooses how to define and treat “tasks” within those guidelines.

ACKNOWLEDGEMENTS

The authors wish to thank Stephanie Binder for initiating the effort for which this task was undertaken, as well as Riley Garrott and Elizabeth Mazzae for their guidance and technical input in shaping this task.

In addition, we gratefully acknowledge the participation of the experts who were interviewed.

Finally, we thank Kim Shelton for her assistance with logistics related to the interviewing process, including teleconference line and transcription arrangements.

REFERENCES

- Alliance Driver Focus-Telematics Working Group. (2006). *Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems*. Washington, D.C.: Alliance of Automobile Manufacturers.
- Angell, L. S., Perez, M., & Hankey, J. (2008). *Driver usage patterns for secondary information systems*. Presented at the First Human Factors Symposium on Naturalistic Driving Methods & Analyses. Blacksburg, VA: Virginia Tech Transportation Institute. August 27, 2008.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- European Union. (2007). *Statement of Principles on HMI. Safe and efficient in-vehicle information and communication systems: Update of the European Statement of Principles on Human Machine Interface*. Brussels, Belgium: Technical Committee of the European Communities.
- Fleischman, E. A. (1975). *Manual for ability requirement scales (MARS), Form A (Original Version)*. Potomac, MD: Management Research Institute.
- Fleischman, E. A. (1991). *Manual for ability requirement scales (MARS), Form A (Revised Version)*. Potomac, MD: Management Research Institute.
- Japan Automobile Manufacturers Association. (2004). *Guideline for In-Vehicle Display Systems, Version 3.0*, Tokyo, Japan. Retrieved from: www.jama.or.jp/safe/guideline/pdf/jama_guideline_v30_en.pdf.
- Jonassen, D. H., Hannum, W. H., & Tessmer, M. (1989). *Handbook of task analysis procedures*. Westport, CT: Praeger.
- Perez, M. A., Angell, L. S., & Hankey, J. (2010). Chapter 8: Naturalistic Driving: Crash and Near Crash Surrogate Distraction Metrics. In Gary Rupp, (Ed.), *Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety*. Warrendale, PA: SAE International.
- Shepard, A. (2001). *Hierarchical Task Analysis*. London: Taylor and Francis.
- Society of Automotive Engineers. (2004). *Navigation and route guidance function accessibility while driving*. (SAE Recommended Practice J2364). Warrendale, PA: Author.
- Zachary, W. W., Ryder, J. M., & Hicinbothom, J. H. (1998/2000). Cognitive Task Analysis and Modeling of Decision Making in Complex Environments. Chapter 13 in Janis A. Cannon-Bowers & Eduardo Salas (Eds.), *Making decisions under stress: Implications for individual and team training*. Washington, DC: American Psychological Association. xxiii, 447 pp. doi: [10.1037/10278-000](https://doi.org/10.1037/10278-000)

Appendix A. Introductory e-Mail for Initial Contact With Experts

Dear [*Name of Contact Person*],

The National Highway Traffic Safety Administration (NHTSA) of the U.S. Dept. of Transportation has initiated a research project with us, the Virginia Tech Transportation Institute, which is intended to explore certain issues related to the development of guidelines for limiting distraction from in-vehicle devices. As a key part of this project, we are speaking with researchers in the field of distraction and in-vehicle device development to get their input, and we would like to include your perspective.

The areas to be discussed will include:

- Current definitions of a “task interaction” (and associated supporting definitions) used in guidelines (e.g., in the guidelines developed by the Alliance of Automobile Manufacturers).
- Ideas for improvements to these definitions, based on new research.
- Ideas for commonizing the interpretation and application of these definitions when used in testing devices.
- Task taxonomies that could be useful for structuring findings on task effects and/or for predicting task effects for new and novel tasks (that may be similar to a category of tasks).

If you are willing to speak with us to discuss these topics, please detach the RSVP form which accompanies this invitation, fill it out, and return it to us by e-mail within 10 days so that we may set up an appointment with you. In order to answer any questions that you may have about this effort, and in order to complete scheduling, we may follow-up with you by phone within the next week, after you have had a chance to reflect on this and return the attached short form.

Participation in these interviews is entirely voluntary.

For your planning purposes, you will receive a short packet prior to the telephone interview (which includes four questions to which you are asked to respond prior to the interview). It will take about 15 minutes to fill out. We ask you to have it with you when we hold the telephone interview. The telephone discussion itself will last between one and two hours -- and will be recorded (if you grant permission) -- otherwise, we will only take written notes of your answers. The intent of the interview is to gather your perspective on these important topics, and we assure you that we will protect the confidentiality and security of the information you provide. We will summarize your answers with those of other experts, in such a way that no particular answer can be traced to any particular expert by anyone outside the research team. Only the VTTI portion of the research team will have access to the interview recordings and notes. The NHTSA portion of the research team will have access to aggregated data (pooled and summarized across interviews). During the interview, we will ask you if you would like to be identified as a contributor to the research effort in any report or publication that should emanate from this work (with the provision that your name will not be associated with any particular response). You will be listed as a contributor only if you wish to be.

Participation in these interviews is entirely voluntary. You are free to opt out, to decline to answer any question at any time, and to withdraw from the interview at any time, without any penalty or consequence. This research study has been reviewed by the Virginia Tech Institutional

Review Board. For research-related problems or questions regarding your rights as a research participant, you may contact Dr. David Moore, Chair, Virginia Tech Institutional Review Board for the Protection of Human Subjects, telephone [redacted]; e-mail: moored@vt.edu; address: Office of Research Compliance, 2000 Kraft Drive, Suite 2000 (0497), Blacksburg, VA 24060. For other questions, please feel free to contact Linda Angell, Derek Viita, or Miguel Perez. Thank you for your help with this important effort; we look forward to speaking with you soon!

Sincerely,

Miguel Perez, Ph.D. & Linda S. Angell, Ph.D.

Human Factors Research Scientists
Center for Automotive Safety Research
Virginia Tech Transportation Institute
Blacksburg, VA 24061
[redacted] -----[redacted]

Appendix B. Script for Telephone Interviews With Experts

CONTRACT NUMBER: DTNH22-05-D-01019

TASK ORDER: TO0033 – Visual-Manual Guidelines

Subtask 2a: Exploring Task Definitions and Task Taxonomies

Script for Telephone Interviews with Experts

Hello – This is [names here] from the Virginia Tech Transportation Institute. Thanks for agreeing to speak with us today. We'd like to set the stage for our discussion today by reminding you that the National Highway Traffic Safety Administration (NHTSA) of the U.S. Dept. of Transportation has initiated a research project with us to explore issues surrounding how "tasks" should be defined under guidelines for limiting and testing distraction from use of devices with visual-manual interfaces. As a key part of this project, we are speaking with researchers in the field to get their input, and we would like to include your perspective. The areas we will discuss are:

- Current definitions of a "task interaction" (and associated supporting definitions) used in guidelines (e.g., in the guidelines developed by the Alliance of Automobile Manufacturers).
- Ideas for improvements to these definitions, based on new research.
- Ideas for commonizing the interpretation and application of these definitions when used in testing devices and systems.
- Task taxonomies that could be useful for structuring findings on task effects and/or for predicting task effects for new and novel tasks (that may be similar to a category of tasks).

Before we get started, we would like to remind you that:

[Circle whichever selection, "a" or "b," that the participant made in his/her initial RSVP to participate and say that one over the phone:

- a. "We will be recording this conversation just to make sure we catch everything. All recordings will be treated as privileged information, stored securely, and shared with no one outside the research team." OR
- b. "We will not be recording this conversation in accordance with your wishes, so we will be taking notes as we discuss things with you. These notes will be treated as privileged information, stored securely, and shared with no one outside the research team."]

We expect this conversation to take between one and two hours.

If you feel uncomfortable answering any or all of the questions, you may opt out (or decline to answer) any questions at any time.

This research study has been reviewed by the Institutional Review Board at Virginia Tech. For research-related problems or questions regarding your rights as a research participant, you can contact: Dr. David Moore, Chair, Virginia Tech Institutional Review Board for the Protection of Human Subjects, telephone [redacted]; e-mail: moored@vt.edu; address: Office of Research Compliance, 2000 Kraft Drive, Suite 2000 (0497), Blacksburg, VA 24060.

Now we would like to begin the interview.

I. BACKGROUND OF EXPERT

- A. Please tell me a little bit about your experience and expertise in working in the area of driver distraction. [Probe for current job responsibilities, educational background/degrees, years of experience.]
- B. (If not mentioned above) Do you also have experience and expertise in **working specifically** on topics related to guidelines or tests related to distraction from use of devices with visual-manual interfaces?
 - i. *No*
 - ii. Yes [If not mentioned in a.] What programs have you directly worked on that involved guideline or test development in the area of driver distraction (or a closely related area)?

II. DEFINITION OF “TASK” IN CURRENT USE

Here is a definition of “TASK” that is often used in testing interactions with devices. “A sequence of control operations leading to a goal at which the driver will normally persist until the goal is reached.”

It is from the Alliance Guidelines (and derived from SAE)

- A. From your perspective, how good is this definition?
- B. Is it the same as -- or similar to -- what you use in your organization? (If not identical, what definition does your organization use?)
- C. Does this definition need improvement? (If yes, what leads you to conclude this?)
- D. What improvements would you suggest?
- E. Do all organizations like yours (who use this type of definition) interpret it like you do and apply it similarly?
- F. If not, what differences are you aware of in how tasks are defined between organizations?

G. What gives rise to these differences?

H. What could be done to minimize these differences and establish common practice throughout the industry/discipline in applying definitions?

III. DEFINITIONS THAT ACCOMPANY THE DEFINITION OF “TASK”

A whole set of definitions have been set forth in the Alliance Driver Focus-Telematics document, which are intended to facilitate common practices in testing adherence of devices and tasks to driver workload limits. These definitions are shown in the table which appears in Addendum A of the *Interview Preparation Packet* that was e-mailed to you. They include definitions for the terms: system function, goal, subgoal, subtask, dependent task, start state, end state, and transitions between tasks. (The definition of “task” is also included in the table for comparison purposes).

Because this is a fairly comprehensive set of definitions that have been developed for use together as a set (during testing of tasks relative to criteria in guidelines), we are asking you to consider this set for our next discussion.

Do you have your e-mail packet handy for reference?

- Yes/No
- When “yes” – Okay, great! Let’s proceed.

Table 7. Definitions From the Alliance Driver Focus-Telematics Guidelines¹

Term	Definition	Special Notes
System Function	A system function consists of a major grouping of related tasks and is defined to be a broad category of activity performed by a system (for example, Navigation). Other examples would be: telecommunication-based services, Internet services, etc. (p.64)	
Goal	A goal is defined as a system state sought by a driver. Driver goals can be met through different system executions and, as such, the system states that correspond to the meeting of these driver goals can be observed and recognized (regardless of the system being operated). That is, goal achievement is defined as equivalent to achieving the system state that meets the driver's intended state, independent of the particular system being executed or method of execution. Examples given (p.64)	<p>NOTE 1: It may be necessary to operationalize a task's end state for evaluation purposes (see "End State" definition).</p> <p>NOTE 2: Clarification regarding multiple ways to reach a goal: When a system provides redundant controls or multiple software-driven paths for a user to reach a goal, all design-intended paths to reach a goal should meet the principles and criteria with representative, typical tasks. Examples given (p. 64)</p>
Subgoal	A subgoal is an <i>intermediate</i> state on the path to the goal toward which a driver is working. It is often distinguishable from a goal in two ways: (1) It is usually not a state at which the driver would be satisfied stopping, and (2) It may vary in its characteristics and/or ordering with other subgoals across hardware/interface executions and thus is system-dependent .	
Task	A task is defined as a sequence of control operations (i.e., a specific method) leading to a goal at which the driver will normally persist until the goal is reached. An example is obtaining guidance from a navigation system by entering a street address using the scrolling list method until route guidance is initiated.	NOTE: this definition only applies to secondary tasks not to the primary driving task.
Subtask	A subtask is defined as a sub-sequence of control operations that is part of a larger task sequence and which leads to a subgoal that represents an <i>intermediate</i> state in the path to the larger goal toward which the driver is working.	
Dependent Task	There is a class of tasks (called "dependent tasks") which may be distinguished from subtasks yet cannot be initiated until another task is first completed. Their "start state" is thus dependent upon the "end state" of another (antecedent) task. However, such tasks are to be treated as tasks (rather than as subtasks) for purposes of evaluating compliance of tasks with the principles and criteria below. They can be distinguished from subtasks by examining their end state (or goal state), which will usually be a driver-sought, system-independent state. Example: After choosing a restaurant from a POI list, the driver is offered an internet function option of making a reservation at the restaurant. The dependent task of making a reservation can be initiated	<p>NOTE 1: Subtasks should not be treated as separate, dependent tasks. For example, entering the street name as part of the navigation destination entry is not a separate task from entering the street number; rather, these are subtasks of the same task.</p> <p>NOTE 2: The concept of "dependent tasks," however, does have special relevance for some domains such as that of route following using a route guidance</p>

¹ Alliance Driver Focus-Telematics Working Group (2006). Page numbers refer to the location of these definitions and examples in the Alliance Guidelines.

	only following the task of selecting a restaurant within the NAV function. It is therefore a separate, dependent task (p. 65)	support system. In particular, after the way-finding mode has been initiated (and destination entered), subsequent route-following guidance can be treated as a series of dependent tasks. For example, following the guidance from point of issuance (“Next turn is on Maple, ¼ mile ahead”)* through achievement of goal (e.g., making of the instructed turn) would be defined as a dependent task whose start state depends on having completed the prior route maneuver successfully. (Such tasks may be evaluated analytically or through engineering judgment, as appropriate.) (p. 65) *Example added for clarity.
Start State of Task	The start of a task under test is the default start screen for the system function under which the task falls. Every task within a system function must share the same start state for purposes of evaluation for compliance with these principles and criteria. An exception is made for tasks that can be initiated only following the completion of a previous task. For these dependent tasks, the start screen would be the end of the previous task (p. 65).	
End State of Task	For the purpose of testing to the criteria contained in this section, the end state of a task is the completion of the final manual input to achieve the driver’s goal, or as indicated by the test subject, as appropriate to accurately measure the duration of the task. This operational definition of task end state is necessary due to the fact that test systems may need to be used for evaluations (outside of a functioning vehicle and outside of functioning network connectivity). As a result, the end state for a task is operationalized to be the completion of control inputs for the task sequence, or as indicated by the test subject, as appropriate to accurately measure the duration of the task. Example: A destination entry task ends with the final control input that initiates way-finding. This is an example of a task that ends with the final control input (p. 65).	
Transitions Between Tasks	One source of workload in a driver’s interactions with an advanced information system is making transitions between tasks in different parts of the system (e.g., moving from navigation functions to radio functions). As such, for purposes of evaluating compliance with the principles and criteria below, transitions between major system functions (e.g., power-up default screen, NAV, phone, Internet, radio, etc.) should be evaluated and, when evaluated, could be treated as separate “tasks.” This method for determining which transitions to evaluate should help identify transitions that have a high expectation, real-world likelihood of consumer use (p. 66).	Example: At system start-up, the telematics display default screen shows the audio system (<i>the top-level screen for the audio system function</i>). When evaluating a NAV task such as destination entry, one must first evaluate the “transition task” of initiating NAV, starting at the audio system display; then one must evaluate the NAV task of destination entry starting with the first NAV display upon function initiation (p. 66).

- I. From your perspective, how good are these definitions **as a set**? Are you aware of any other set of definitions that covers all (or nearly all) of these concepts or terms? (If yes, probe for source and citations).
- J. Is this **set** of definitions one that you use in your organization or is it similar to what your organization uses? (If not identical, which definition/s in the set does your organization use or which ones are different?)
- K. Are there any improvements to the definitions in this set that are needed? Are all the definitions clear? Are any of them confusing or ambiguous?
- L. What improvements would you suggest? What leads you to suggest each of these? (e.g., new research or issues that surfaced while using the existing ones or . . . probe)
- M. Do all organizations like yours (who use these definitions) interpret them like you do and apply them similarly?
- N. Would you say that industry practice (and practice in the human factors discipline) in applying these definitions is fairly consistent/common/uniform? Would you say that a device and task would be tested similarly, regardless of which organization was doing the testing?
- O. If not, what differences are you aware of in how tasks are defined between organizations?
- P. What gives rise to these differences? Can the differences be addressed through improved definitions? Are there certain definitions that are vulnerable to misinterpretation or to misapplication?
- Q. What could be done to minimize these differences and establish common practice in applying definitions? (Would it help to provide examples of how to apply these definitions?)
- R. Focusing specifically on the definition for “dependent tasks” (as opposed to “subtasks”) is this definition clear enough? Can it be applied in a consistent way across organizations? Do you think it is? (Probe).
- S. And focusing specifically on the definition for “start state,” is this definition clear? Can it be applied in a consistent way across organizations? Do you think it is? (Probe).

IV. APPLYING THE CURRENT DEFINITIONS

In order to be sure that we understand how you are using the definitions, we provided some material to you through e-mail that explored how your organization defines “start” and “end” states of tasks for testing purposes.

We would now like to discuss these items with you. Do you have them handy?
Okay. Great!

The pages we will use for this discussion are on page 2, 3, 4, and 5 of your packet.
Could you tell us what your answers are for:

	TASK START	TASK END	
Part A:	_____	_____	(page 2 of packet)
Part B:	_____	_____	(page 3 of packet)
Part C:	_____	_____	(page 4 of packet)
Part D:	_____	_____	(page 5 of packet)

Thank you!

Do you have any questions or comments that you would like to make about defining
“start” or “end” states of tasks?

Okay, then we will go on to the next part of the interview.

V. EXPLORING ALTERNATIVE DEFINITIONS

- A. Now we would like you to consider some *alternative* definitions for the concept of a “TASK.”
- B. We would like you to rate whether you think each of these definitions is an improvement or not over the first definition you saw. **We’d like you to make your ratings using a three-point scale of: better, same, or worse.** (This is shown in Addendum B of your packet, if you wish to look at it.)
- C. We will read you an alternative definition and then ask for your rating. There are three alternative definitions all together.

	Worse	Same	Better
a. Alternative Definition 1 for “Task”	_____	_____	_____
“A sequence of operations leading to a goal at which the driver will normally persist until the goal is reached . . . noting that these operations consist of any and all responses of the driver to the task (e.g., glances, as well as control inputs, etc.). “Task start” begins with the first activity on the task, and “task end” finishes with the final response to the task. These initial and end responses may be glances rather than control inputs. Note: To capture all responses, the “task start” during testing may be operationally defined as beginning as soon as the tester finishes saying, “Please begin now.” The final observable response to the task may be the driver saying, “I’m done” in a test.			

b. Alternative Definition 2

“A sequence of secondary task activities leading to a pause in the behavioral record during which the driver re-directs his/her attention back to the roadway and to driving for a period of time long enough to re-establish situation awareness (which typically takes 1 to 3 seconds, or perhaps more, depending on conditions).”

c. Alternative Definition 3

“A sequence of operations leading to a goal at which the driver will normally persist until the goal is reached . . . or until a driver need is satisfied.” (This is intended to accommodate “natural” tasks, in which drivers may persist in an activity to satisfy a personal need . . . such as searching and searching for something good to listen to on a long drive, for example).

Okay, thank you!

VI. TAXONOMIES

Now we would like to switch focus away from definitions of terms. We would like to ask you to think next about taxonomies of tasks – structures for organizing types of tasks, based on similarities between them – as well as for organizing data about them (both test data as well as perhaps empirical data from the scientific literature).

A. Does your organization use a taxonomy of tasks for any of the following purposes?

As I read each item, *please let me know (by saying yes or no) whether your organization uses a taxonomy for each application that I read:*

- i. Organizing data previously collected on tasks?
- ii. Performing engineering analysis of new tasks early in design?
- iii. Determining which new tasks may need to be empirically tested?
- iv. Grouping tasks and selecting a representative task from the group to be tested (to streamline testing)?
- v. Applying redlines (or criteria) to tasks that are similar to ones in the taxonomy that have been previously tested?

B. If your organization does use a taxonomy, can you describe its structure and attributes?

- i. Would you recommend its use to others?
- ii. Are there any features that you would improve about it?
- iii. Are you aware of any (other) taxonomies in use – or in the literature – that would be appropriate for achieving the purpose for which you are using a taxonomy?

- C. If your organization does NOT use task taxonomies, why not? What are the reasons?
- i. Has your organization found that it has no need for them?
 - ii. Or has it found that there is not a satisfactory taxonomy available?
 - iii. Or some other reason?

VI. TYPES OF TAXONOMIES

Next, we would like you to consider **two types of taxonomies**.

Each of these could be populated with test results accumulated over time – but for purposes of illustration they have been left empty for now.

Please take a moment to review each of the taxonomies which were in your e-mailed *Interview Preparation Packet in Addendum C and D (shown on the next two pages for the interviewer)*, and then when you are ready, we would like to ask you to rate how useful each taxonomy might be for the testing program already conducted in your organization (or perhaps to-be-conducted in your organization).

Do you have the material handy?

Okay, great! To make your ratings, we would like to ask you to use a rating scale that is shown in *Addendum E of your e-mail packet*. It has a **midpoint of zero, a positive side, and a negative side**. On this scale, **the midpoint of zero** would represent a taxonomy that offers no added value. Then on the positive side of the scale, ratings go up to +4 at the far end, with a rating of +4 representing that this type of taxonomy would be “extremely useful and beneficial.” (Lower positive numbers would represent lesser degrees of usefulness and benefit, but all positive numbers indicate some degree of utility and benefit.) However, the negative side of the scale indicates inappropriateness, disbenefit, or concerns. On the negative side of the scale, ratings go down to -4, and a rating of -4 would indicate that this type of taxonomy would be extremely inappropriate for use and a source of concern or disbenefit. (Negative values between 0 and -4 would represent lower degrees of inappropriateness, concern, or disbenefit.)

Any questions?

Rating scale:

+4 Extremely useful and beneficial

+3

+2

+1

0 No Added Value

-1

-2

-3

-4 Extremely inappropriate for use; a source of concern; a source of disbenefit

Are you ready to rate the two types of SAMPLE taxonomies?

_____ RATING FOR the first TAXONOMY (Task Attributes - Alt. A)

Ask: What led you to give it that rating? (Probe)

_____ RATING FOR the second TAXONOMY (Task Complexity - Alt. B)

Ask: What led you to give it that rating? (Probe)

(NOTE: If any ratings below zero are given, particularly probe for concerns.)

Alternative 1: Taxonomy Based on Task Attributes

Task Type	Display	Control	Pages	TSOT	Total Glance Time to Task (85% ile of sample)	Mean Glance Duration during Task (85% percentile of sample)
Search for Item using Menu-Based Paging	Organized list with X line items per screen (meeting separate Guidelines)	Soft buttons labeled on screen	Y pages allowed in total			
Access lower level function through visual-manual inputs starting at system level screen	Organized display of system functions on screen (meeting separate guidelines)	Labeled soft buttons and hard buttons for high frequency actions (Number of buttons limited to z in number)	Menu depth limited to X			

Alternative 2. Taxonomy Based on Task Complexity and Naturalistic Data

Complexity Level	Task Properties	Specific Tasks in Interface <i>Double Asterisk ** identifies in-vehicle device task</i>	Probable Types of Users	Est. Frequency of Use	Est. Conditions of Use	TSOT	Total Glance Time to Task (85 th percentile)	Mean Duration of Single Glances (85 th percentile)	Estimated Crash Risk (Odds Ratio)
High	Five or More Button Presses or Steps	Enter destination in nav system (using keypad)** Dial handheld device ** Reach for moving object Insect in vehicle Read while driving Apply makeup							
Moderate	Three to Four Button Presses or Steps	<i>Adjust HVAC mode, temp, fan**</i> <i>Insert/retrieve CD**</i> <i>Talk/listen to handheld device**</i> Eat in vehicle Look at exterior object Reach for non-moving object Other personal hygiene							

Low	One to Two Button Presses or Steps	Activate Embedded Telematics System (turn on)** Engage cruise control** Passenger in adjacent seat Comb hair							
-----	------------------------------------	---	--	--	--	--	--	--	--

VII. CLOSING QUESTIONS

Is there anything that you would like to add to the discussion that we’ve had today? That is, is there some topic or issue that we did not touch on during the interview that you feel might be valuable?

Okay. . . thank you! That completes the interview.

May we just ask you one final thing: Would you like to have your name included as a contributor to this research in any report or publication that may emanate from this work? Please note that if we were to list your name, it would not be identified with any particular response or responses; rather, the list of contributors would be separated from the aggregated data and provided as part of an acknowledgement.

RESPONSE: YES: ____ NO: ____

Thank you *so very much* for your time and your contributions!

Appendix C. Interview Preparation Packet

To: Interviewee

This packet is sent to you in order to help you prepare for your scheduled interview. There is no need to return it to us. However, prior to your scheduled interview, we do ask that you please read and respond to the first four pages of this packet. We will ask you for your responses during our interview discussion with you.

In making those responses, you are welcome to use Addendum A for reference, if you like – and to confer with colleagues in your organization.

The remaining pages of this packet are for your reference and use during the interview itself (they show rating scales that will be used). Please make sure this packet is handy and accessible to you during your scheduled interview.

Thank you!! We look forward to talking with you soon!

Linda Angell, VTTI Project Scientist
Phone: [redacted]

Applying the Set of Task-Related Definitions to Task Analyses

PART A. Below you will find an analysis of a task, decomposed into a numbered sequence of steps. Using the set of definitions from the Alliance Guidelines, please determine where the “TASK START” and “TASK END” would be demarcated for purposes of testing performance on this task *in your organization*. Please indicate your answer by filling in the blanks below, which are labeled *Task Start* and *Task End*. In the first blank provided below, enter the first step that should be included in the measurement of the task (the starting step to be included in measurement), and in the second blank, enter the last step in the task that should be included prior to its end (the final step to be included in the measurement of the task). If it is helpful, you may draw a line through the box to indicate *Task Start* and then draw a second horizontal line through the box to indicate where the *Task End* should be demarcated. You are welcome to refer to the definitions in the Addendum as you respond to this part of the questionnaire (they are provided for your convenience), and you are welcome to consult with other members employed by your organization. We will ask you to report these values in your phone interview.

TASK START: _____ **TASK END:** _____

TASK A	
Goal:	Enter destination of 1040 South Elm Street, Greenlee, Indiana
Task:	<ol style="list-style-type: none"> 1. Look for and find the system function button (a soft button) which brings up “NAVIGATION” functions 2. Press that system function button to bring up the main “NAVIGATION” screen 3. Press the (soft) button which allows for “Address Entry” 4. Press (soft) button for entry by “Street and House Number” 5. At prompt, enter name of state and city <ol style="list-style-type: none"> a. Either use keypad to enter ‘state’ using letter-by- letter entry -- and then ‘city’ using letter-by- letter entry, OR b. Select state from list at top of screen by touching it, and then selecting city from the next list that appears by touching it. 6. At prompt, enter name of street <ol style="list-style-type: none"> a. Either use keypad to enter name of street letter-by- letter, OR b. Select street from list at top of screen by touching it 7. At prompt, enter address <ol style="list-style-type: none"> a. Either use keypad to enter numbers of address, OR b. Select address from list at top of screen by touching it 8. Push “GO” so the map will display and route instructions will begin. 9. Look at map and receive first instruction.
Device:	Embedded navigation system with large color screen near top of center stack

PART B. Again, please do the same thing that you did in Part A, only on a new task, TASK B, shown below.

We will ask you to report these values in your phone interview.

TASK START: _____ **TASK END:** _____

TASK B
Goal: Place call to person at 751-9367
Task: <ol style="list-style-type: none">1. Move hand toward phone, using eyes to guide its position2. Push "Phone" to select phone functions3. Select "dialer"4. Recall phone number5. Move hand and eyes to phone keypad6. Enter digit7. Enter digit8. Enter digit9. Enter digit10. Enter digit11. Enter digit12. Enter digit13. Push «send»
Device: Hand-held phone docked in vehicle with earbud in place prior to task

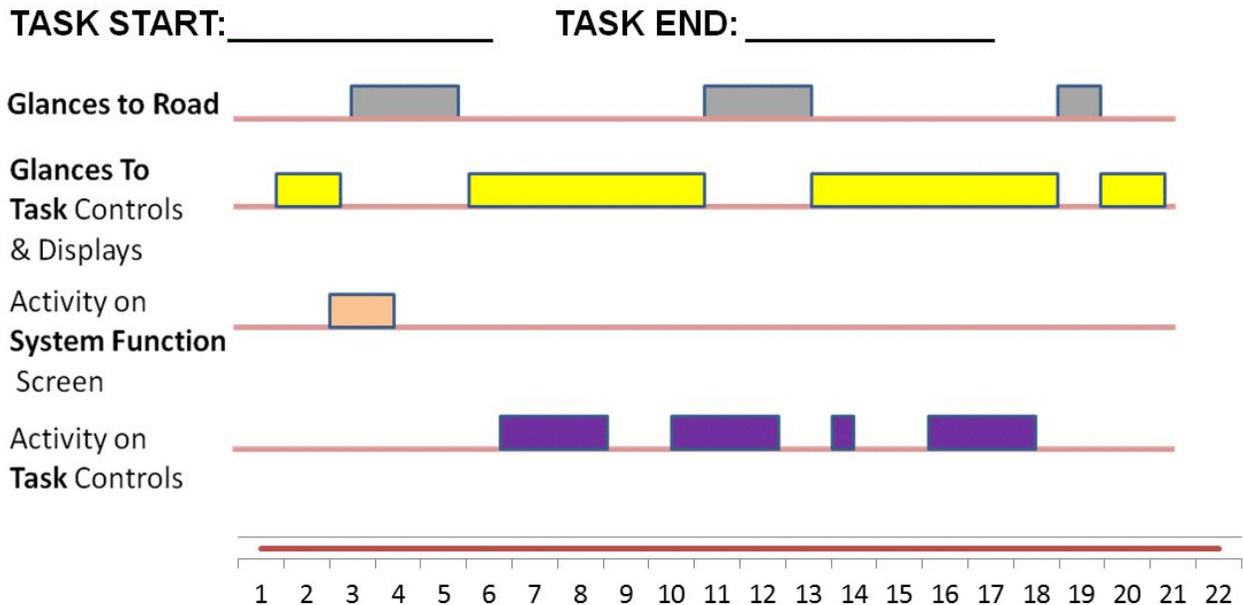
Using the Behavioral Test Record in Conjunction With the Set of Task-Related Definitions

PART C. This time, we would like you to do something similar, but instead of using a task analysis to represent a task, we would like you to make your determinations based on a record of driver behavior in the task. So below you will find a time history of responses to a particular task, with each horizontal line in the record illustrating a different type of response (glances, button pushes, etc.) over time (from left to right).

We would *still* like you to use the set of definitions from the Alliance Guidelines, and we would still ask that you please determine where the “TASK START” and “TASK END” would be demarcated within this test of performance data *in your organization*.

Please indicate your answer by filling in the blanks below, which are labeled *Task Start* and *Task End*. Use the “ruler” under the figure to help fill in the blanks.

If it is helpful, you may draw a vertical line through the figure to indicate *Task Start* and then draw a second vertical line through the figure to indicate where the *Task End* should be demarcated. At the point where your vertical “task start” line intersects the ruler, take a reading from the ruler at that point and enter that value into the *Task Start* blank. At the point where your vertical “task end” line intersects the ruler, take a reading from the ruler at that point and enter that value into the *Task End* blank. You are welcome to refer to the definitions in the Addendum as you respond to this part of the questionnaire (they are provided for your convenience), and you are welcome to consult with other members employed by your organization. We will ask you to report these values in your phone interview.



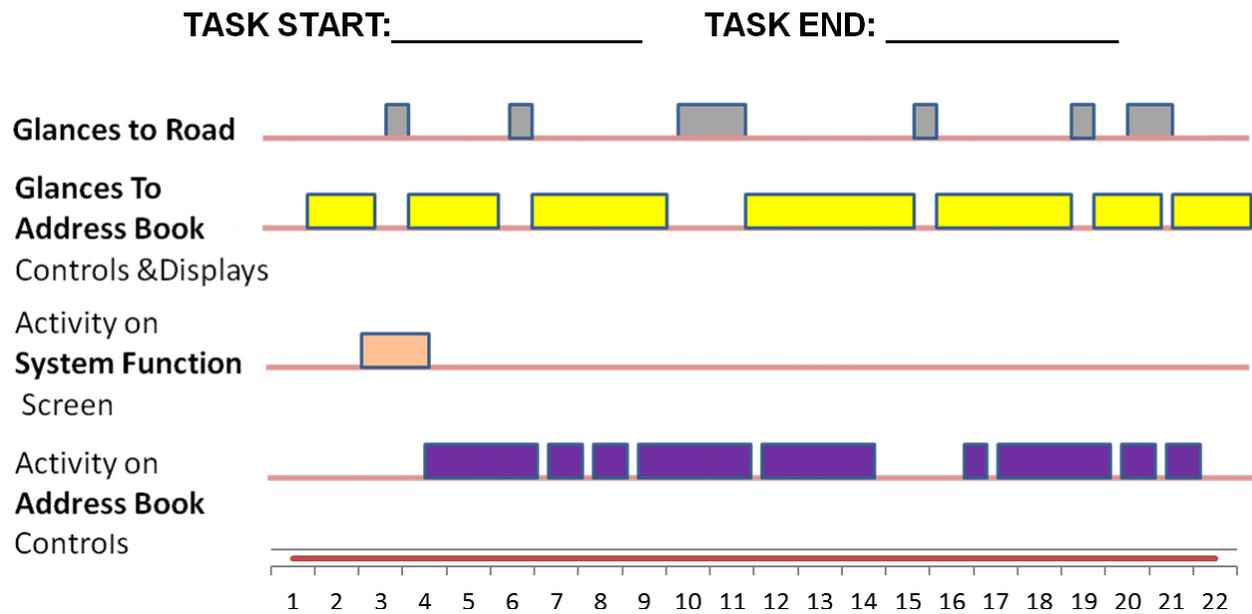
Using the Behavioral Test Record in Conjunction With the Set of Task-Related Definitions

PART D. This time, we would like you to do the same thing as in the prior example, only we would like you to consider a different task (the one shown below). We have repeated the instructions, for your reference.

Below you will find a time history of responses to a particular task, with each horizontal line in the record illustrating a different type of response (glances, button pushes, etc.) over time (from left to right). We would *still* like you to use the set of definitions from the Alliance Guidelines, and we would still ask that you please determine where the “TASK START” and “TASK END” would be demarcated within this test of performance data *in your organization*.

Please indicate your answer by filling in the blanks below, which are labeled *Task Start* and *Task End*. Use the “ruler” under the figure to help fill in the blanks.

If it is helpful, you may draw a vertical line through the figure to indicate *Task Start* and then draw a second vertical line through the figure to indicate where the *Task End* should be demarcated. At the point where your vertical “task start” line intersects the ruler, take a reading from the ruler at that point and enter that value into the **TASK START** blank. At the point where your vertical “task end” line intersects the ruler, take a reading from the ruler at that point and enter that value into the **TASK END** blank. You are welcome to refer to the definitions in the Addendum as you respond to this part of the questionnaire (they are provided for your convenience), and you are welcome to consult with other members employed by your organization. We will ask you to report these values in your phone interview.



Addendum A

Table 8. Definitions From the Alliance Driver Focus-Telematics Guidelines²

Term	Definition	Special Notes
System Function	A system function consists of a major grouping of related tasks and is defined to be a broad category of activity performed by a system (for example, navigation). Other examples would be: telecommunication-based services, Internet services, etc. (p.64)	
Goal	A goal is defined as a system state sought by a driver. Driver goals can be met through different system executions and, as such, the system states that correspond to the meeting of these driver goals can be observed and recognized (regardless of the system being operated). That is, goal achievement is defined as equivalent to achieving the system state that meets the driver's intended state, independent of the particular system being executed or method of execution. Examples given (p.64)	<p>NOTE 1: It may be necessary to operationalize a task's end state for evaluation purposes (see "End State" definition).</p> <p>NOTE 2: Clarification regarding multiple ways to reach a goal: When a system provides redundant controls or multiple software-driven paths for a user to reach a goal, all design-intended paths to reach a goal should meet the principles and criteria with representative, typical tasks. Examples given (p. 64)</p>
Subgoal	A subgoal is an <i>intermediate</i> state on the path to the goal toward which a driver is working. It is often distinguishable from a goal in two ways: (1) It is usually not a state at which the driver would be satisfied stopping, and (2) It may vary in its characteristics and/or ordering with other subgoals across hardware/interface executions and thus is system-dependent .	
Task	A task is defined as a sequence of control operations (i.e., a specific method) leading to a goal at which the driver will normally persist until the goal is reached. An example is obtaining guidance from a navigation system by entering a street address using the scrolling list method until route guidance is initiated.	NOTE: this definition only applies to secondary tasks not to the primary driving task.
Subtask	A subtask is defined as a sub-sequence of control operations that is part of a larger task sequence and which leads to a subgoal that represents an <i>intermediate</i> state in the path to the larger goal toward which the driver is working.	
Dependent Task	There is a class of tasks (called "dependent tasks") which may be distinguished from subtasks yet cannot be initiated until another task is first completed. Their "start state" is thus dependent upon the "end state" of another (antecedent) task. However, such tasks are to be treated as tasks (rather than as subtasks) for purposes of evaluating compliance of tasks with the principles and criteria below. They can be distinguished from subtasks by examining their end state	NOTE 1: Subtasks should not be treated as separate, dependent tasks. For example, entering the street name as part of the navigation destination entry is not a separate task from entering the street number; rather, these are subtasks of the same task.

² Alliance Driver Focus-Telematics Working Group (2006). Page numbers refer to the location of these definitions and examples in the Alliance Guidelines.

	<p>(or goal state), which will usually be a driver-sought, system-independent state. Example: After choosing a restaurant from a POI list, the driver is offered an internet function option of making a reservation at the restaurant. The dependent task of making a reservation can be initiated only following the task of selecting a restaurant within the NAV function. It is therefore a separate, dependent task (p. 65)</p>	<p>NOTE 2: The concept of “dependent tasks,” however, does have special relevance for some domains such as that of route following using a route guidance support system. In particular, after the way-finding mode has been initiated (and destination entered), subsequent route-following guidance can be treated as a series of dependent tasks. For example, following the guidance from point of issuance (“Next turn is on Maple, ¼ mile ahead”)* through achievement of goal (e.g., making of the instructed turn) would be defined as a dependent task whose start state depends on having completed the prior route maneuver successfully. (Such tasks may be evaluated analytically or through engineering judgment, as appropriate.) (p. 65) *Example added for clarity.</p>
Start State of Task	<p>The start of a task under test is the default start screen for the system function under which the task falls. Every task within a system function must share the same start state for purposes of evaluation for compliance with these principles and criteria. An exception is made for tasks that can be initiated only following the completion of a previous task. For these dependent tasks, the start screen would be the end of the previous task (p. 65).</p>	
End State of Task	<p>For the purpose of testing to the criteria contained in this section, the end state of a task is the completion of the final manual input to achieve the driver’s goal, or as indicated by the test subject, as appropriate to accurately measure the duration of the task. This operational definition of task end state is necessary due to the fact that test systems may need to be used for evaluations (outside of a functioning vehicle and outside of functioning network connectivity). As a result, the end state for a task is operationalized to be the completion of control inputs for the task sequence, or as indicated by the test subject, as appropriate to accurately measure the duration of the task. Example: A destination entry task ends with the final control input that initiates way-finding. This is an example of a task that ends with the final control input (p. 65).</p>	

<p>Transitions Between Tasks</p>	<p>One source of workload in a driver’s interactions with an advanced information system is making transitions between tasks in different parts of the system (e.g., moving from navigation functions to radio functions). As such, for purposes of evaluating compliance with the principles and criteria below, transitions between major system functions (e.g., power-up default screen, NAV, phone, Internet, radio, etc.) should be evaluated and, when evaluated, could be treated as separate “tasks.” This method for determining which transitions to evaluate should help identify transitions that have a high expectation, real-world likelihood of consumer use (p. 66).</p>	<p>Example: At system start-up, the telematics display default screen shows the audio system (<i>the top-level screen for the audio system function</i>). When evaluating a NAV task such as destination entry, one must first evaluate the “transition task” of initiating NAV, starting at the audio system display; then one must evaluate the NAV task of destination entry starting with the first NAV display upon function initiation (p. 66).</p>
--	--	---

Addendum B

Rating Scale #1 To Be Used During Interview (Definitions)

Worse Same Better

— — —

Addendum C

Alternative 1: Taxonomy Based on Task Attributes

Task Type	Display	Control	Pages	TSOT	Total Glance Time to Task (85%ile of sample)	Mean Duration Of Single Glances during Task (85th percentile)
Search for Item using Menu-Based Paging	Organized list with Xline items per screen (meeting separate guidelines)	Soft buttons labeled on screen	Y pages allowed in total			
Access lower level function through visual manual inputs starting at system level screen	Organized display of system functions onscreen (meeting separate guidelines)	Labeled soft buttons and hard buttons for high frequency actions (Number of buttons limited to z in number)	Menu depth limited to X			

Addendum D

Alternative 2. Taxonomy Based on Task Complexity and Naturalistic Data

Complexity Level	Task Properties	Specific Tasks in Interface <i>Double Asterisk ** identifies in-vehicle device task</i>	Probable Types of Users	Est. Frequency of Use	Est. Conditions of Use	TSOT	Total Glance Time to Task (85 th percentile)	Mean Duration of Single Glances (85 th percentile)	Estimated Crash Risk (Odds Ratio)
High	Five or More Button Presses or Steps	Enter destination in nav system (using keypad)** Dial handheld device ** Reach for moving object Insect in vehicle Read while driving Apply makeup							
Moderate	Three to Four Button Presses or Steps	<i>Adjust HVAC mode, temp, fan**</i> <i>Insert/retrieve CD**</i> <i>Talk/listen to handheld device**</i> Eat in vehicle Look at exterior object Reach for non-moving object Other personal hygiene							

Low	One to Two Button Presses or Steps	Activate Embedded Telematics System (turn on)** Engage cruise control** Passenger in adjacent seat Comb hair							
-----	------------------------------------	---	--	--	--	--	--	--	--

Addendum E

Rating Scale #2 To Be Used During Interview (Taxonomies)

- +4 Extremely Useful and Beneficial
- +3
- +2
- +1
- 0 No Added Value
- 1
- 2
- 3
- 4 Extremely inappropriate for use; a source of concern; a source of disbenefits

DOT HS 811 782
May 2013



U.S. Department
of Transportation
**National Highway
Traffic Safety
Administration**



8704-051713-v2