SAE International™

SURFACE VEHICLE INFORMATION REPORT

SAE J2830 JUL2008

Issued 2008-07

Process for Comprehension Testing of In-Vehicle Icons

RATIONALE

Icons and symbols can be used to communicate information to the driver in a manner that is not dependent on language and can save valuable space on in-vehicle displays. Incomprehensible icons, however, have the potential to negatively affect safety (e.g., if the driver does not understand the icon). Despite the ubiquity of icons and symbols within the in-vehicle environment, few guidelines exist for testing of icons and symbols. Key shortcomings of existing icon testing procedures include: a lack of contextual information provided to experimental subjects, over-reliance on evaluator judgment regarding how well an experimental subject comprehended an icon's meaning, and a lack of prescriptive information—feedback to icon designers regarding how individual icons could be improved based on the results of comprehension testing.

INTRODUCTION

This information report provides a valid and reliable process for comprehension testing of candidate automotive icons and symbols. The comprehension testing process was developed though a multi-year effort, supported by the SAE International (SAE) and other organizations, aimed at generating a test methodology that would: yield high-quality comprehension data for new automotive symbols, provide clear and specific guidance back to symbol developers based on the test results, and could be adopted and performed internationally to support international standards efforts.

TABLE OF CONTENTS

TO PLACE A DOCUMENT ORDER: **Tel: 877-606-7323 (inside USA and Canada)**
Tel: 724-776-4970 (outside USA)
Fax: 724-776-0790
Email: CustomerService@sae.org
SAE WEB ADDRESS: http://www.sae.org

1. SCOPE

This document describes a process for testing the comprehension of symbols or icons. Although the process may be used to test any symbols or icons, it has been developed specifically for testing ITS active safety symbols or icons (e.g., collision avoidance), or other symbols or icons that reflect some in-vehicle ITS message or function (e.g., navigation, motorist services, infotainment). Within the process, well-defined criteria are used to identify the extent to which the perceived meaning matches the intended meaning for a representative sample of drivers. Though the process described below reflects a paper-and-pencil approach to conducting the testing, electronic means (i.e., conducted using a computer) can be used as well. The data or results from this process are analyzed to assess the drivers’ comprehension of the symbol or icon. These data will be used to provide guidance in the design of in-vehicle symbols or icons.

2. REFERENCES

2.1 Applicable Publications

The following publications form a part of this specification to the extent specified herein. Unless otherwise specified, the latest issue of SAE publications shall apply.

2.1.1 SAE Publications

Available from SAE International, 400 Commonwealth Drive, Warrendale, PA 15096-0001, Tel: 877-606-7323 (inside USA and Canada) or 724-776-4970 (outside USA), www.sae.org.

Campbell, J. L., Kludt, K., and Kiefer, R. J. (2007). *Evaluation of In-vehicle Symbols for an Intersection Crash Avoidance System*. Paper presented at the 14th Asia Pacific Automotive Engineering Conference (APAC-14), August 5-8, 2007, Hollywood, CA

Campbell, J. L., Hoffmeister, D.H., Kiefer, R. J., Selke, D. J., Green, P. & Richman, J. B. (2004). Comprehension Testing of Active Safety Symbols (SAE Technical Paper No. 2004-01-0450). *Proceedings of the SAE 2004 World Congress*. Warrendale, PA: SAE International. (Also published by SAE in SP-1877, *Human Factors in Driving and Telematics, and Seating Comfort*)

Saunby, C. S., Farber, E. I., & DeMello, J. (1988). Driver Understanding and Recognition of Automotive ISO Symbols. *SAE Technical Paper Series* (No. 880056). Warrendale, PA: SAE International

2.2 Related Publications

The following publications are provided for information purposes only and are not a required part of this document.

ANSI Z535 3 (American National Standards Institute). (1998). *American national standard criteria for safety symbols.* Washington, DC: National Electrical Manufacturers Association

Balbale, U., & Kiefer, R. J. (June, 2001*). Driver preference rankings of various collision avoidance-related symbols*. Presented at ISO TC22/SC13/WG5 (Road Vehicles-Symbols for Controls, Indicators and Telltales) meeting in Dresden, Germany

Campbell, J. L., Carney, C., Richman, J. B., & Lee, J. D. (2004). *In-vehicle display icons and other information elements. Task F: Final in-vehicle symbol guidelines* (Final Report). Seattle, WA: Battelle Human Factors Transportation Center. (see also http://www.tfhrc.gov/safety/pubs/03065/index.htm)

Collins, B. L. (1982). *The development and evaluation of effective symbol signs*. Washington, DC: U.S. Department of Commerce (NBS Building Science Series, 141)

Green, P. (1993). Design and evaluation of symbols for automobile controls and displays. In B. Peacock & W. Karwowski (Eds.), *Automotive Ergonomics* (chapter 12, pp 237- 268). London, UK: Taylor and Francis

Green, P. (1981). Displays for automotive instrument panels: Production and rating of symbols. *HSRI Research Review, July-August*, 1-12

Green, P., & Pew, R. W. (1978). Evaluating pictographic symbols: An automotive application. *Human Factors*, *20*(1), 103-114

ISO/DIS 9186. (1988). *Procedures for the development and testing of public information symbols*. Geneva, Switzerland: International Standards Organization

Richman, J. B., Campbell, J. L., & McCallum, M. C. (2002). Effective IVIS comprehension research: Context and response scaling methods [CD-ROM]. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society*, 1939-1943

Wolff, J. S., & Wogalter, M. S. (1998). Comprehension of pictorial symbols: Effect of context and test method. *Human Factors, 40*(2), 173-186

3. PROCESS STEPS

3.1 Step 1: Review Candidate Icons for Comprehension Testing

3.1.1 Candidate icons are obtained for testing.

Icons should be solicited from individuals and organizations interested in conducting comprehension testing for candidate icons. All icons should be submitted as electronic graphic files, and should be as similar as possible, in terms of size, appropriate color, resolution, etc., to the future production application.

3.1.1.1 Pre-testing

Prior to comprehension testing, it is recommended that all icons undergo some pre-testing, such as Production Tests or Appropriateness Ranking Tests (see also Step 3 below for procedures to incorporate an Appropriateness Ranking Test into the comprehension testing).

3.1.1.2 Message

All icons should have a message, as well as a general in-vehicle function (i.e., navigation, collision avoidance, motorist services, etc.) assigned to them. The message assigned to each icon should be carefully selected by the organization submitting the icon(s), as this message will be compared to subjects' responses when determining overall comprehension rates.

Also the organization submitting the icon(s) should, as appropriate, identify any "major" vs. "minor" components of the message to support the data analyses conducted in Step 4. "Major" elements are those words or phrases within the definition or message associated with an icon that are critical to the meaning of the icon. "Minor" elements are those words or phrases within the definition or message associated with an icon that are less critical. For example, for a "Tire pressure warning-low" icon, the "tire" and "warning" parts of the message are "major" elements, while the "pressure" and the "low" parts of the message are "minor". The logic here is that the icon's primary purpose is to warn of a tire problem and its secondary purpose is to describe the exact nature of the problem.

In this regard, it is very important that the message (or definition) identified for each icon be as clear and complete as possible in order to support accurate and efficient scoring of the subjects' responses in Step 4 of this procedure. Be aware that even small changes to the intended message can lead to large differences in the final comprehension scores for individual icons. If appropriate, comprehension responses from the subjects can be scored against more than one definition.

3.1.1.3 Context

The individuals/organizations submitting the icon(s) for testing should also provide suggestions for the context descriptions given to subjects during testing (see also Appendix A).

3.1.1.4 Priorities for comprehension testing: Active safety icons should have highest priority for testing.

3.1.2 For each study, 10-20 candidate icons should be included in the comprehension testing.

These 10-20 icons will likely reflect a number of distinct messages, not just one or two.

3.2 Step 2: Prepare for Comprehension Testing

3.2.1 Prepare test materials.

Assemble test materials as follows:

a. Prepare icons so that they are as similar as possible, in terms of size, appropriate color, resolution, etc., to the future production application.

b. Place candidate icons on separate sheets of paper, slides, or computer screens, depending on the presentation method planned for the study.

One presentation approach that can work well when testing icons that have already been introduced into a vehicle fleet is to use photographs as the test stimuli, showing the icon-to-be-tested in its real-world context. Although using photographs can make the test preparation process more difficult, it improves the level of realism associated with the comprehension testing and increases the validity of the test. To the extent possible, the photographs should show the icon co-located with functionally similar icons, as well as any associated controls. An example of this presentation approach is shown below in Figure 1.

FIGURE 1 - EXAMPLE OF USING A PHOTOGRAPH AS TEST STIMULI, SHOWING THE ICON-TO-BE-TESTED, AS WELL AS SURROUNDING AREAS OF THE VEHICLE INTERIOR

c. In addition to the candidate icons tested, some standard icons should be tested in order to provide comparison data. Approximately 10-15% of the total icons tested should be common, standard in-vehicle icons (such as seat belt, oil pressure, and temperature icons).

d. Randomize presentation order across subjects.

d. Separate different candidates for the same message into distinct test sets.

f. Prepare and provide to subjects an example sheet with a common icon (like a fuel pump to indicate a fuel gage) and its meaning written beneath the graphic. Figure 2 provides an example of this.

g. Develop context statements per Appendix A (see attached).

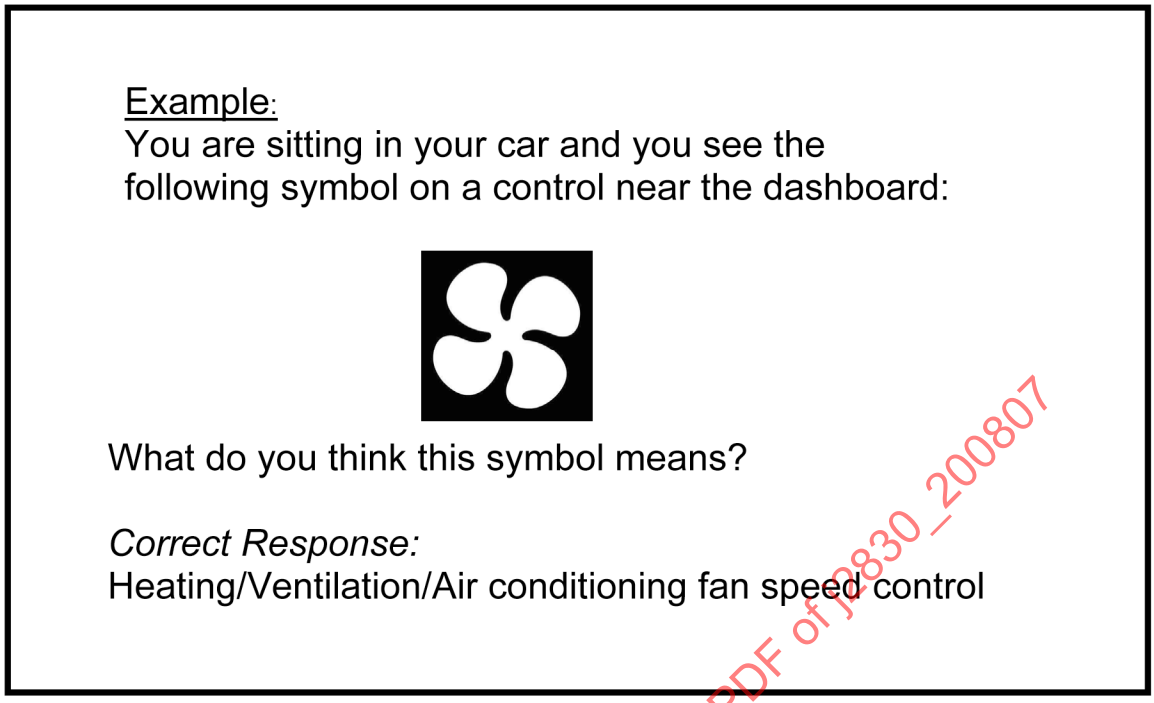

FIGURE 2 - EXAMPLE OF A COMMON ICON THAT CAN BE PRESENTED TO SUBJECTS PRIOR TO TESTING

3.2.2 Control for learning effects.

When multiple candidate icons for the same message share some common design elements and look very similar to one another, there is the possibility of learning effects associated with testing these icons. That is, as subjects are asked to identify a series of icons that look similar to one another, their ability to do so is improved across trials. Counterbalancing the presentation order of these icons (or randomizing the presentation order), can have the effect of distributing any learning effects, but does not remove them. Another approach is to split up similar icons into two (or more) separate booklets, in which individual subjects are only tested with a single booklet (all non-similar icons can remain in both booklets). This approach has the disadvantage of essentially doubling the number of subjects needed for the testing, but removes learning effects as a possible rationale for varying comprehension levels across similar-looking icons.

3.2.3 Schedule test subjects.

Use the following criteria when scheduling test sessions:

a. The total number of subjects should be 30-40.

b. All subjects should be licensed drivers who drive at least twice a month.

c. The goal for subject population is an approximately equal mix of older/younger, male/female subjects.

d. In general, subjects groups should be as representative as possible of the larger driving population; diverse backgrounds and fields of study are preferred.

e. Testing is typically done with groups of 10-20 subjects in order to increase the overall efficiency of individual sessions.

For example, an adequate number of younger subjects can be obtained from local universities. Past studies in this area have placed an advertisement in the daily newspaper at a local university or on *craigslist*, an on-line classifieds service. The advertisement provides a very general description of the research being conducted, as well as a telephone number to call for those who are interested in participating. Individuals responding to the ads can be screened and, as appropriate, scheduled for a data collection session. Older subjects can be obtained in a similar fashion using local newspapers instead of the university newspaper. It may be more efficient to directly contact local retirement centers or senior centers. Using this method, an initial telephone call can be made to key administrative staff at the retirement or senior center, followed, perhaps, by a direct presentation to the potential subjects themselves describing the purpose of the study. Once adequate interest and availability of subjects at the retirement or senior center have been established, an on-site data collection session can be scheduled.

3.3 Step 3: Conduct Comprehension Testing

3.3.1 Provide subjects with instructions and examples.

a. Indicate the context in which the icon will be used (see Appendix A).

b. Explain the task that subjects will perform during the testing.

3.3.2 Test candidate symbols with a representative group of subjects.

Present test subjects with candidate icons and ask them to write down the action, condition, activity, location, etc. that they believe is represented by the icon.

3.3.3 If not conducted prior to the comprehension testing, conduct appropriateness ranking testing.[1]

After the comprehension testing is complete, ask subjects to rank order the candidate symbols for a particular message in terms of the degree to which they think each symbol communicates the message, in the following manner:

a. On a separate page/section of the response booklet, state the true and complete meaning for the symbols, and present the symbols vertically in a column.

b. Ask the subjects to place a “1” next to the symbol that they think is the best symbol for communicating the message, a “2” next to the symbol that they think is the second best symbol for communicating the symbol, and so on, until all the candidate symbols have been ranked according to their perceived appropriateness.

3.4 Step 4: Analyze Comprehension Data and Summarize Results

3.4.1 Analyze data.

3.4.1.1 Train at least two scorers. Have at least two (2) trained scorers categorize responses along a scale according to well-defined criteria that identify the likelihood that an individual response indicates correct comprehension of the icon. Adequately trained scorers should, at least: (1) be familiar and comfortable with the criteria presented in Table 1, (2) practice scoring sample data from a representative comprehension test or a pre-test, and (3) have a chance to compare and reconcile these practice scoring efforts against the results of other scorers. The “major” and “minor” components of each icon’s message (as provided by the organization submitting the icon—see Step 1 above) should be used to categorize the subjects’ responses. That is, the perceived meaning (the subject’s response) should be compared to the intended meaning (provided by the organization submitting the icon).

[1] It was noted above that—typically—production tests and appropriateness ranking tests should be conducted prior to conducting comprehension testing. However, appropriateness ranking data can be a useful complement to comprehension data, and can be collected and scored with very little effort. In particular, the appropriateness ranking test can be a useful way to resolve “ties” between candidate icons that result from the comprehension testing.

3.4.1.2 Score responses for each icon. Table 1 should be used to make this comparison between perceived and intended meaning and to categorize the subject's responses. For each icon candidate, convert the total number of responses in each category into percentages.

During the scoring process, scorers should maintain a detailed score sheet of the specific subject responses assigned to each of the nine (9) comprehension score categories for each icon tested. These can be provided back to the organization that originally submitted the icons for comprehension testing. Appendix B provides an actual example of this score sheet (taken from recent icon comprehension tests), as well as a blank score sheet.

In order to support the evaluation of the comprehension results, as well as subsequent design decisions arising from the testing, it may also be helpful to note the frequency of individual responses on the score sheet.

3.4.1.3 Get scorer consensus. Once all of the subjects' responses for an individual icon have been scored by the two trained scorers, the scorers should compare their scores, reconcile any differences, and develop a single "consensus" set of scores for each icon. These consensus scores will represent the final scores for each icon.

3.4.1.4 Diagnose responses. Of the nine (9) scoring categories in Table 1, scores of 1 or 2 are the most important, as they—when summed—define overall comprehension rates. The remaining scores (particularly 3, 4, 5, and 9) are most useful for diagnostic purposes; i.e., identifying problems with candidate icons and possible methods for improving comprehension.

For safety-critical icons, identify the number and percentage of critical confusions or errors (Category 9 in Table 1). Critical confusions or errors reflect responses that indicate that the subject perceived the message to convey a potentially unsafe action. For non-safety-critical icons, such errors will probably fall into Category 5 or 6.

3.4.2 If appropriateness ranking data were collected, analyze these data.

Calculate average rankings for each symbol; lower average rankings correspond to higher levels of perceived appropriateness.

TABLE 1 - RATING SCALES FOR CATEGORIZING AND SCORING SUBJECT RESPONSES TO THE ICONS

Comprehension Score	Description
1	The response matches the intended meaning of the icon exactly.
2	The response captures all major informational elements of the intended meaning of the icon, but is missing one or more minor informational elements.
3	The response captures some of the intended meaning of the icon, but it is missing one or more major informational elements.
4	The response does not match the intended meaning of the icon, but it captures some major or minor informational elements.
5	The response does not match the intended meaning of the icon, but it is somewhat relevant.
6	Participant's response is in no way relevant to the intended meaning of the icon.
7	Participant indicated he/she did not understand the icon.
8	No answer.
9	For safety-critical icons, identify the number and percentage of critical confusions or errors. Critical confusions or errors reflect responses that indicate that the participant perceived the message to convey a potentially unsafe action.

Table 2 provides sample responses for the “Tire pressure warning-low” icon (shown in Figure 3) and their corresponding ratings, taken from our recent icon comprehension research.

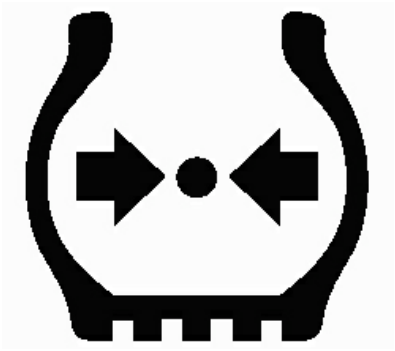

FIGURE 3 - THE “TIRE PRESSURE WARNING-LOW” ICON

TABLE 2 - SAMPLE RESPONSES AND RATINGS FOR A “TIRE PRESSURE WARNING-LOW” ICON

Subject Response	Rating	Description
Warning - low tire pressure	1	The response matches the intended meaning of the icon exactly.
Tire pressure needs attention	2	The response captures all major informational elements of the intended meaning of the icon, but it is missing one or more minor informational elements.
Normal tire pressure	3	The response captures some of the intended meaning of the icon, but it is missing one or more major informational elements.
You have a mechanical problem	4	The response does not match the intended meaning of the icon, but it captures some major or minor informational elements.
Tire symbol	5	The response does not match the intended meaning of the icon, but it is somewhat relevant.
Be aware of what is ahead	6	Participant's response is in no way relevant to the intended meaning of the icon.

3.4.3 Interpret and summarize results.

Decisions regarding criteria for minimum percent correct comprehension rates for individual icons should reflect designer’s needs, as well as the consequences associated with selecting a cutoff that is too high or too low. ISO requirements for an acceptable symbol have been a (minimum) 66 percent correct comprehension level (i.e., combined categories 1 and 2 from Table 1), while ANSI specifies that 85% correct comprehension should be used. ANSI specifies that more than 5% critical confusions (with a sample of at least 50 participants) should result in rejection of the icon.

Also, the distribution of responses across the nine (9) score categories should be taken into consideration. For example, consider a situation where Candidate Icon A obtains the following distribution of rating scores across the 9 categories: 1-2: 30%, 3-4: 5%, 5-8: 65%, 9: 0% and Candidate Icon B obtains the following distribution of rating scores across the 9 categories: 1-2: 30%, 3-4: 65%, 5-8: 5%, 9: 0%. Although the “correct comprehension” rate (combined scores of rating score categories 1 and 2) for the two icons is the same at 30%, it is clear that comprehension of Icon B was better, and that Icon B will likely need less revision in order to increase the number of subject responses that fall into either Category 1 or Category 2 from Table 1.

Based on the data analyses, provide the detailed scoring sheets and any available diagnostic information (which could include specific recommendations on how individual icons could be revised to improve comprehension) back to the individuals/organizations that submitted the icons.

3.5 Comparison of SAE and ISO Methods

This general scoring approach has been successfully used to assess comprehension for hundreds of in-vehicle icons in past studies. Although it requires some initial training and takes slightly longer to score, it provides some advantages over the simpler 7-point scales used by ISO.

Both the proposed SAE scoring approach and the ISO scoring approach result in an overall comprehension score for icons that have been tested. However, a key advantage of the proposed SAE approach over the ISO approach is that the proposed SAE approach makes distinctions between the major and minor elements of the message, and gives greater weight to major elements. While this distinction is not relevant to all icons (for example, icons with simple, one-word messages), the ISO approach does not allow such distinctions. Overall, the proposed SAE approach allows the comprehension testing team to provide better, more specific diagnostic information back to the original icon designers because it allows for analysis of what element(s) of the icon the subjects seemed to understand, and what element(s) may have led to errors in comprehension.

4. NOTES

4.1 Marginal Indicia

A change bar (I) located in the left margin is for the convenience of the user in locating areas where technical revisions, not editorial changes, have been made to the previous issue of this document. An (R) symbol to the left of the document title indicates a complete revision of the document, including technical revisions. Change bars and (R) are not used in original publications, nor in documents that contain editorial changes only.

PREPARED BY THE SAE SYMBOLS SUBCOMMITTEE
OF THE SAE SAFETY AND HUMAN FACTORS COMMITTEE

APPENDIX A - PROVIDING SUBJECTS WITH CONTEXT DURING ICON EVALUATIONS

A.1 INTRODUCTION

This appendix provides some guidance and examples of ways to provide appropriate context to experimental subjects during comprehension evaluations of candidate icons. In real-world driving, icons are presented in the *context* of certain in-vehicle capabilities and driving circumstances. As such, evaluations of in-vehicle icons should include a description of the context in which the icons will be presented and used.

Icon evaluations should avoid providing either too little or too much context to experimental subjects. If too little context is provided, unrealistically low comprehension scores may result because the subjects may be unable to connect a visual icon with the many possible icon meanings. If too much context is provided, unrealistically high comprehension scores may result because the subjects have been cued for a certain response by the specificity of the context. Both these extremes should be avoided.

The context provided to subjects should describe the: (1) general capabilities of the in-vehicle information system (IVIS) or technology that will be used to present the icons and (2) general driving circumstances associated with the presentation of the icon by the in-vehicle system.

A.2 PRESENTATION OF GENERAL SYSTEM CAPABILITIES

The general IVIS capabilities will likely reflect the range of in-vehicle functions or messages presented by the entire set of icons that are being tested. If only collision avoidance icons are being tested, then only a general description of the collision avoidance system is needed. For example, an evaluation of a broad set of in-vehicle icons might provide the following “system capabilities” information to subjects:

> Our purpose is to investigate issues related to the use of graphical icons in the in-vehicle environment. Recent advances in automotive technology have allowed the development of various devices that can present a wide range of trip, vehicle, and safety information to drivers. Much of this information will be provided to drivers through color displays located on the instrument cluster or center console of the vehicle. These display systems might present information such as:

Adaptive Cruise Control: An improved cruise control system that automatically maintains adequate spacing between the driver’s car and the vehicle ahead.

Augmented Signage: System that displays regulatory (for example, speed limit) and road condition icons (for example, winding road) for the road on which the driver is presently driving. The displayed icons would be similar to the signs commonly posted along the roadside.

Collision Avoidance System: On-board sensor and display systems that warn the driver of an impending collision with another vehicle or object, a vehicle lane departure, and other motion-based safety hazards.

Motorist Services: System that provides an in-vehicle “Yellow Pages” function for drivers. This directory could present information regarding the availability and locations of motorist services such as gas stations, restaurants, hotels, or recreational activities.

Trip Navigation: System that provides the driver with route assistance and trip planning information. This includes driving directions, cautions about certain routes, current traffic conditions, road construction, or weather updates.

Vehicle Condition Monitoring: Device that informs the driver of problems with the vehicle or vehicle systems (oil, tire pressure, engine, etc.).

Subjects should also be provided with a simple schematic drawing of a vehicle interior, supplemented with arrows or shaded regions to show them where the icons might be displayed.