RESEARCH ON NETWORK-DISTRIBUTED
VOICE-ACTIVATED SYSTEM ARCHITECTURE
FOR TELEMATICS SERVICES

Makoto Shioya, Senior Researcher
Hitachi Systems Development Laboratory (HSDL)
292 Yoshida-cho, Totsuka-ku, Yokohama-shi, Kanagawa 244-0817, Japan
Tel: +81-45-860-2428, Fax: +81-45-860-1675
E-mail: shioya@sdl.hitachi.co.jp, Web: http://www.sdl.hitachi.co.jp/

Nobuo Hataoka, Chief Researcher
Hitachi Central Research Laboratory (HCRL)
1-280 Higashi-Koigakubo, Kokubunji-shi, Tokyo 185-8601, Japan
Tel: +81-42-323-1111, Fax: +81-44-959-0851
E-mail: hataoka@crl.hitachi.co.jp, Web: http://koigakubo.hitachi.co.jp/

Takuya Nishimoto, Research Associate
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
Tel: +81-3-5841-6902, Fax: +81-3-5841-6902
E-mail: nishi@hil.t.u-tokyo.ac.jp, Web: http://hil.t.u-tokyo.ac.jp/

Juhei Takahashi, General Manager
Takashi Odajima, Project Manager
Research Dept.
Association of Electronic Technology for Automobile Traffic and Driving (JSK)
25-5 Toranomon 1-chome, Minato-ku, Tokyo 105-0001, Japan
Tel: +81-3-3519-4396, Fax: +81-3-3501-5685
E-mail: takahas@jari.or.jp, todajima@jari.or.jp, Web: http://www.jari.or.jp/

SUMMARY

This paper describes the construction of a voice-activated system architecture for telematics services. The main features include: (a) a network-distributed voice-activated system, (b) HMI dialogue control based on emergency and priority considerations and (c) conduciveness to standardization of voice commands, effect sounds and dialogues. The effectiveness of this architecture has been confirmed in actual driving tests that include situations where the driver's workload increases, such as when passing or making right-hand turns.

INTRODUCTION

Background

Visual Distraction and Mind Distraction
A voice-activated human-machine interface (HMI) is an effective way of reducing the driver's workload and visual distraction when operating an onboard information system to access the Internet and obtain telematics services. However, it has been pointed out that even hands-free and voice-activated HMI technology may not always be safe to use depending on the nature of the service involved or the
driving state of the vehicle at the time. This is believed to be caused by the mind distraction compared to the visual distraction (Figure 1).

**HMI Classes and Situation-Aware HMI Control**

It started the discussions to meet distractions which take various degrees according to the driving abilities of individuals and the situations. As a result, an introduction of the HMI classes and the situation-aware HMI control were considered (Figure 1).

**Load Share between In-Vehicle System and Network**

Moreover, assuming that telematics services provide real-time information, it will be necessary for the in-vehicle system and the network to share the voice processing function and the processing load (Figure 2).

**ITS Information Tags for Priority Control**

On the other hand, the discussions have been held whether to give the ITS information tags to the information sent to a vehicle so that the system could discriminate the priority of the information for the vehicle (Figure 1).

**Objectives**

Network-Distributed Voice-Activated System Architecture

To meet these requirements, we have constructed the architecture of a voice-activated system that incorporates network-distributed processing capability. The architecture adopts dialogue control system which takes into account driving situation, emergency and priority to reduce mind distraction. The tests have been undertaken to verify the effectiveness of the system architecture.

Study of Guidelines for Voice-Activated System

We also have conducted a study of the related guidelines, concerning with voice commands, effect sounds or music, timing and dialogues. The dialogues have been written using the standard language ‘Voice XML’ and the interface to the driver has been constructed according to the standard ‘SALT’-like event driven style.

**Fig.1 Background and Objective**

![Diagram showing the relationship between Autonomous or Dedicated Telematics Car Navigation Systems, Network-Distributed Telematics Systems, HMI Guideline for Telematics Systems, Easing of Restrictions on Traffic Information, ITS Information Tag, and Network-Distributed Telematics System to meet Safety & Situation.](attachment:image.png)
ARCHITECTURE OF VOICE-ACTIVATED SYSTEM FOR TELEMATICS SERVICES

System Architecture

The architecture of the voice-activated system consists of an in-vehicle system, a center system and application service providers as shown in the Figure 2. Each constituent system incorporates an engine for voice recognition and synthesis and an HMI server function.

Fig.2 Concept of Network-Distributed Voice-Activated Telematics Service System

Center System, In-Vehicle System, and HMI Server

The rough architecture of the center system and the in-vehicle system is shown in the Figure 3. The ASPs are at the similar positions as the center system is in the figure. In some cases, the driver’s desired information can be provided from the database of the in-vehicle information system. In other cases, such as when a driver desires information about parking availability at that time, the HMI server transfers control to the voice-activated engine and HMI of the center system. In other words, basically, (1) static or fixed information related dialogues are processed by the in-vehicle system, (2) dynamic or real-time changing information related dialogues are processed by the center system or the ASPs, (3) local or detailed information related dialogues are processed by the center system or the ASPs. This is the load sharing function.

In addition, the HMI server constantly monitors the situation in and around the vehicle, the driver’s state, the state of an HMI dialogue and whether there is information from an external source, so as to sense whether any distraction tend to occur, any high priority external information come in, and so on. If distraction tendency or high priority external information is sensed, the HMI server changes the nature of the dialogue according to the circumstances, if necessary. These are the distraction sensing function and the dialogue control function.
Driving & Environmental Conditions

In-Vehicle System

Request
- Distraction Sensing
- Dialogue Control considering Info. Priority & Driver’s Skill
- Load Share btw Center / In-Vehicle

HMI Server
- Sensing
- Control
- Share

Driver’s Conditions

Vehicle

Center System

Communications

Networks

Fig.3 Voice Telematics HMI System Architecture

HMI Dialogue Control

The dialogues are written using the standard language ‘Voice XML’ and are decomposed basically into small sub-dialogues, each of them consists of a prompt and an answer, so that the situation-aware dialogue control to reduce the possibility of mind distraction is easy to realize. The sub-dialogues are picked up and combined into the series of dialogues for the telematics service, taking into account the situations at that time. In other words, (1) Service dialogues consist of the series of sub-dialogues. (2) One sub-dialogue basically consists of the prompt from the system and the responding speech of a driver.

The HMI server changes the combination of the sub-dialogues flexibly and immediately when the situation changes.

If the dialogues do not proceed well by some reason, the HMI server selects prepared dialogues appropriate to the situation and guides the driver to get out the situation. For examples:
(3) If there is no responding speech to the prompt or the system can not understand the responding speech, the sub-dialogues are combined and executed in the next order.
   (a) Repeats the prompt,
   (b) Executes the command ‘HELP’,
   (c) Repeats the same prompt as (a),
   (d) Presents the recommended responding speech and confirm the driver’s agreement,
   (e) Executes the command ‘HOW to USE’,
   (f) Executes the command ‘STOP’, to close the dialogue and return to the main menu.

The repetition times are adjustable.

Based on this architecture, the situation-aware dialogue control has been realized by preparing various kinds of sub-dialogues which are elemental parts described by the Voice XML and the dialogue transition table under the various situations.

HMI Server Architecture and Control Process

The HMI server architecture rather in detail is shown in the Figure 4. The HMI dialogue control which considers the driving conditions, the emergency and the priority of information is consisted of the following processes.
(1) Interrupt information is received at the situation manager and the possibility of a mind distraction caused by the situation change is judged by the situation manager.
(2) The priority manager judges the priority of the information taking into account the driving conditions, and it rearranges the processes in the priority order on the waiting list at the process manager.
(3) The process manager controls the processes of pause dialogue, resume dialogue, etc., according to the waiting list.
(4) The scenario manager matches the appropriate scenarios for the user, and has the process manager to execute them via the dialogue manage table.
(5) The dialogue histories are saved at the process manager, to provide for the case when a dialogue pauses, jumps to the other scenario and backs to resume at the previous scenario.
(6) The standard command processor for backward positioning in the case of the dialogue resume is prepared at the process manager.

The control is executed in combination of dialogue pause, dialogue resume, scenario change, scenario ordering, interaction timing, interface medias ( sounds, voice, images, etc. ), and so on. The basic interface medias are:
(a) At vehicle speed is zero, visual oriented multi-modal interface ( image, text, voice, effect sound ).
(b) At normal driving workload, voice oriented multi-modal interface ( route guide map, voice, effect sound ).
(c) At high driving workload, pause dialogue at the beginning and resume dialogue at the end (with effect sound).

The control is executed also in consideration of each driver’s skill.

STUDY OF VOICE-ACTIVATED SYSTEM GUIDELINES FOR TELEMATICS SERVICES

A study was made of the following items that should be shared in common so that a driver could dialogue with the system simply and safely while driving, regardless of the vehicle types.
As for developing tools, the standard VoiceXML descriptive language was used in constructing the dialogue function.

**Study of Voice Command Guideline**

Voice commands are to be selected, assuming that telematics services are obtained within a framework that does not interfere with the execution of driving operations (Table 1). In other words, sub task (service access) must not interfere with main task (driving). From this viewpoint, we selected two types of command functions:

(1) Command function that enables drivers to control the progress of a dialogue as they wish. Examples of commands are:
- Start (Open): start dialogue
- Wait (Pause): pause dialogue
- Continue (Resume): resume dialogue
- Jump: jump to coming service step
- Back:: back to past service step
- Slowly: say prompt more slowly
- Fast: say prompt more fast
- Stop (Close): stop dialogue

(2) Command function for adjusting the degree of control over the progress of a dialogue. Examples of commands are:
- Many, More
- a Little, Little by Little
- Reverse

These are used for quick decision of resume position and used together with:
- Back
- Jump
- etc.

These are also used for adjusting speed of dialogue progress and used together with:
- Wait
- Slowly
- Fast
- etc.

**Table 1  A Study of Voice Command Guideline**

| Sub Task (Service Access) Must Not Interfere with Main Task (Driving) |  
| A: Command Function to Control Dialogue Progress according to Driver’s Intention |  
| - Start (Open) |  
| - Wait (Pause), Continue (Resume) |  
| - Jump: Jump to Coming Service Step |  
| - Back: Back to Past Service Step |  
| - Slowly, Fast: Say Prompt More Slowly or Fast |  
| - Stop (Close) |  
| B: Command Function to Adjust Dialogue Progress Control |  
| - Many, More, a Little, Little by Little; Reverse |  
| → For Quick Decision of Resume Position |  
| Use Together with ‘Back’, ‘Jump’, etc. |  
| → For Adjusting Speed of Dialogue Progress |  

**Study of Effect Sound Guideline**

Sound prompts or effect sounds are useful, assuming that a sound prompt, music and a pause will be needed to convey the state of a dialogue (i.e., mode) to drivers without involving unnecessary attention on their part in the process of obtaining a telematics service (Table 2). In other words, sub task (service access) must not give drivers excessive load. Effect sounds, music, and timing are
useful to inform progress to drivers with little load. From this viewpoint, we considered two types of effect sounds:

(1) Effect sounds for reassuring drivers by letting them know the status of the system. Examples of system status are:
- Request Accepted: system accepted driver’s request
- Under Processing: system is processing services (searching, etc.)
  - No Waiting: get result soon
  - Waiting: take more time (proceed to the next step is expected)
  - processing service, reconnecting center.

(2) Effect sounds for letting drivers know what action they should take and the reason. Examples of drivers’ action are:
- Utterance is Prompted: driver is prompted to utter (to say)
  - System is Waiting Input for Prompt
  - Repeat: system could not recognize
  - Other Word: system recognized, but not expected word
- Malfunction: action such as reset, etc. is requested (waiting comes to nothing)
  - System is Malfunctioning
  - Network Connection is Cut

Table 2 A Study of Effect Sound Guideline

<table>
<thead>
<tr>
<th>Task</th>
<th>Must Not Give Driver Excessive Load</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A: Inform System State to Feel Easy</strong></td>
<td>Effect Sound, Music, Timing are Useful to Inform Progress</td>
</tr>
<tr>
<td>- Request Accepted: System Accepted Driver’s Request</td>
<td></td>
</tr>
<tr>
<td>- Under Processing: System is Processing Services (Search, etc.)</td>
<td></td>
</tr>
<tr>
<td>- No Waiting: Get Result Soon</td>
<td></td>
</tr>
<tr>
<td>- Waiting: Take More Time (Proceed to the Next Step is Expected)</td>
<td></td>
</tr>
<tr>
<td>- Processing Service, Reconnecting Center</td>
<td></td>
</tr>
<tr>
<td><strong>B: Driver is Requested to Act Something (with the Reason)</strong></td>
<td></td>
</tr>
<tr>
<td>- Utterance is Prompted: Driver is Prompted to Utter</td>
<td></td>
</tr>
<tr>
<td>- System is Waiting Input for Prompt</td>
<td></td>
</tr>
<tr>
<td>- Repeat: System could not Recognize</td>
<td></td>
</tr>
<tr>
<td>- Other Word: System Recognized, but not Expected Word</td>
<td></td>
</tr>
<tr>
<td>- Malfunction: Action such as Reset, etc. is Requested</td>
<td></td>
</tr>
<tr>
<td>(Waiting Comes to Nothing)</td>
<td></td>
</tr>
<tr>
<td>- Server is Malfunctioning</td>
<td></td>
</tr>
<tr>
<td>- Network Connection is Cut</td>
<td></td>
</tr>
</tbody>
</table>

Study of HMI Dialogue Guideline

HMI dialogue control is necessary to be executed, according to safety, emergency and priority considerations and the driving skill of individual drivers. We considered four basic cases of controlling dialogues (Table 3):

(1) Cases requiring control of a dialogue before a dangerous situation develops. In these cases, the dialogues are paused, effect sounds are inserted and announcements with the reason are inserted. For example, when winker is turned on (in right or left turn, passing, etc.), the system informs kindly to the driver and controls the dialogue carefully.

(2) Cases where a potentially dangerous situation has developed. In these cases, the dialogues are paused quickly, and no effect sound or announcement. For example, in sudden braking, effect sounds or announcements are too much for the driver.
(3) Cases where the vehicle receives high priority information from an external source, such as information concerning traffic restrictions. In these cases, the dialogues just in progress are paused and the dialogues concerning the external high priority information are started. For example, the system informs the driver before the right turn, if the street after the right turn is impassable.

(4) Cases where a danger warning is issued by a system that monitors safe vehicle operation. In these cases, the dialogues are paused quickly and give priority to the safe vehicle operation monitoring system. For an example, give top priority to hazard alarm from a collision alarm system.

Table 3  A Study of HMI Dialogue Guideline Considering Emergency & Priority

A: Control before get into Dangerous Situation
   - Pause Dialogue / Insert Effect Sound / Insert Announcement (Reason)
     ➔ When Winker Turn ON (Right / Left Turn, Passing, etc.), Inform Kindly and Control Dialogue Carefully

B: On Getting into Dangerous Situation
   - Pause Dialogue Quickly
     ➔ Effect Sound and Announcement is Too Much in Sudden Braking

C: On Receiving Useful Information such as Temporal Traffic Regulation
   - Pause Dialogue in Progress and Start Dialogue of External Information
     ➔ Inform before Right Turn if Street after Right Turn is Impassable

D: On Receiving Hazard Alarm from Safety Driving System
   - Pause Dialogue Quickly and Give Priority to Hazard Alarm System
     ➔ Give Top Priority to Hazard Alarm from Collision Alarm System, etc.

VERIFICATION TESTS OF THE EFFECTIVENESS OF THE VOICE-ACTIVATED SYSTEM ARCHITECTURE

Relations between Driving Load, Telematics Load, Driver’s Skill & Driving Ability Remained

The effectiveness of using this system architecture to control HMI dialogues has been verified by the tests on the streets of Gotenba-City. The tests have been conducted in situations where the workload of obtaining a telematics service have been added to the workload of operating a vehicle. The relations between the workloads, two driver’s skills and the driving ability remained to each driver are shown in the Figure 5. The telematics load is assumed to be constant and as examples of the increased driving workload situations, passing, turning right or left and merging with traffic among others are shown in the figure.

Example of Dialogue Control for Driver B

The control of HMI dialogues has been executed so that the total workload has not exceeded the driving abilities of individual drivers. An example of the HMI dialogue control for the driver B is shown in the Figure 6. In this control case, the dialogues are paused when the driving load plus the telematics
load is assumed to exceed the driver B’s skill. The passing and the mixing cases correspond to this condition in the figure.

**Fig.5  Relations between Driving Load, Telematics Load, Driver’s Skill & Driving Ability Remained**

**Fig.6  An Example of Dialogue Control for Driver B**

**Test System**

The test system used for the effectiveness tests of using the architecture consists of the two vehicles, one for the main subject of the tests (the in-vehicle system) and the other for the center system and the ASP. The connection between these two vehicles are made through the wireless LAN (IEEE802.11b) access point. Voices are transformed into the VoIP form using the Microsoft Net Meeting in each vehicle and are transferred to the other vehicle. Rough structure of the test system is shown in the Figure 7.

The HMI devices consists of a color LCD display with a touch panel, a microphone headset and two speakers, one for voices from the in-vehicle system and the other for voices from the center system. A
driver can touch the screen to burge in (cut in) the proceeding dialogue at any time, and can use any voice commands to start a new dialogue.

As for the in-vehicle system, the center system and the ASP, note PCs are used. The test conditions such as driving and environmental conditions, service priority, driver’s conditions are input real-timely by an operator in the main subject vehicle.

**Fig.7 Test System**

**Dialogue Control in Case of Right Turn**

The main test situation is the right or left turn case. When a vehicle approaches an intersection and a driver switches the winker on to turn the intersection right or left, the HMI server senses the increase of the workload for the driver and pauses the proceeding dialogue. When the vehicle turns and the driver switches the winker off, the HMI server resumes the dialogue paused. These processes are illustrated in the Figure 8 in a right turn case.

A dialogue is paused in a turning period basically, but the figure shows a case when the dialogue is resumed before the end of the vehicle turn. This is the case when the vehicle speed is zero and the driver burges in and uses the command ‘Resume’. This function is prepared for realizing flexibility of the HMI control to adapt various situations and individual drivers.

**Fig.8 Dialogue Control in Case of Right Turn**
Test Conditions

Voice Commands

Effect Sounds
Four effect sounds for informing the system state are prepared. They are for ‘utterance prompted’, ‘utterance accepted’, ‘pause / resume’ and ‘high priority Information’.

Announcements
The announcements are used for informing ‘pause / resume’ and ‘high priority information’.

Telematics Services
As for the telematics services, searching and reserving a parking lot and a restaurant are prepared. In searching and reserving a parking lot, detailed information or related information services are available. On the other hand, in searching and reserving a restaurant, they are controlled to be not available for safety when a vehicle is running.

When a vehicle gets the high priority or highly emergent information, the system pauses and interrupts the dialogue for normal telematics services, and informs the emergency information. After that, the dialogue resumes. This control is according to the concept of ITS information tags now under discussion in Japan.

Test Results

Evaluation by Questionnaire Surveys
The tests were performed on the streets of Gotenba-City by over ten subject drivers who were the members of this project or the specialists of vehicles or in-vehicle devices. After the tests, we have the evaluations of the tests shown below by questionnaire surveys to the drivers.

(1) Safety and Convenience:
Every driver agrees that it feels much safe due to the control of dialogues, and also that a certain level of convenience drop is tolerable.

(2) Pause of Dialogues:
Though the evaluations are scattered, most drivers agree the necessity of the effect sounds and the announcements.

(3) Resume of Dialogues:
The evaluations are scattered for all items, such as the timing of resume, the resume position in the previous dialogue, etc..

(4) Voice Commands:
Every driver agrees the necessity of preparing standard commands. Many drivers request more discussion of the contents of the commands.

(5) Effect Sounds and Announcements:
Most drivers agree the necessity and the effectiveness of the effect sounds and the announcements. The standardization is also requested.

Consideration
Throughout the tests, it reveals that the drivers’ degree of experience in the system and the profiles of individuals affect the evaluation of the system. For examples:
(a) The same timing, the same position, and the same informing style of the pause and the resume cause the different evaluations.
(b) The pause or the informing the pause itself makes a driver to be nervous a little in some
The optimum settings for a dialogue control, voice commands, effect sounds, etc. must be done considering these.

In summary, the knowledge we get is:

(1) It is important the balance of the value of the information needed for a driver and the timing of informing it to the driver.

(2) It is needed the discussion of the multi-modal interface consisted of the combination of plain figures and intuitive color usage, besides the voice and the effect sounds.

(3) It is necessary the function to have the system change to adapt a drivers experience degree in the system.

CONCLUSIONS

It reveals that mind-free concept is important even in voice-activated telematics services system, besides so called hands-free and eyes-free concept. We conclude that controlling the voice dialogues in case of high workload on drivers is effective for reducing mind distraction. It become clear that standardized voice commands and dialogue patterns increases usability of the system. But as for the control contents, the evaluations split into the one that most subject drivers agree and the other that many drivers make different evaluations. The reasons are not clear whether that come from degree of experience in the system or from profiles of individuals. This means that the control exists which most drivers consider much safe qualitatively, but the quantitative knowledge of the boundaries is not given. The knowledge for the items for which the evaluations scatter is only that they may be classified into several patterns.

It is desirable for drivers to be able to obtain telematics services safely with HMIs configured on the basis of the same standard regardless of which car model they drive. This research and development project represents only the first step toward the construction of a system that makes effective use of voice activation for obtaining telematics services from the standpoint of safety. Toward that end, the authors work together with various organizations in conducting quantitative tests and in developing guidelines with the aim of creating a common standard.

ACKNOWLEDGEMENTS

This paper is the result of three years research and development project by the Association of Electronic Technology for Automobile Traffic and Driving (JSK), under contract with MSSF, founded by METI. Our special thanks are due to Professor Tetsunori Kobayashi of Waseda University for guiding our work. We also express our thanks for the members of the Working Group and the related organizations and persons. JSK has been reorganized into JARI (Japan Automobile Research Institute ) this July.