ELSEVIER

Displays 23 (2002) 65–73

DISPLAYS

www.elsevier.com/locate/displa

Acoustic User Interface (AUI) for the auditory displays

Hidenobu Takao[a], Kaoru Sakai[b], Jun Osufi[c], Hiroaki Ishii[a]

[a]*Department of Welfare Systems Engineering, Kanagawa Institute of Technology, Atsugi, Kanagawa 243-0292, Japan*
[b]*Design and Planning Department, Design Centre, Fujitsu Ltd., 4-1-1 Kamikodanaka, Nakahara-Ku, Kanagawa 211-8588, Japan*
[c]*Strategic Business Development Department, Planning and Coordination Division, Mobile Entertainment Company, Pioneer Corporation, 4-1 Meguro I-chome, Meguro-ku, Tokyo 153-8654, Japan*

Abstract

When designing a multimodal human–machine interface, the human function allocation must be considered. To reduce the load on the visual sense, the other sensory channels must be utilized. Here, the authors put forward a HMI which provides unique concepts called acoustic user interface (AUI).

With an AUI, information can be removed from the visual channel and presented through the auditory channel. The head-related transfer function is used to create the impression of sounds arriving from any direction, either inside or outside the visual field of view. As well as reducing the loading on the visual sense, aural 'images' can be 'projected' coincident with the location of images displayed on the visual displays, in order to co-ordinate and communicate more effectively multimedia information. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Acoustic user interface; Auditory display; Binaural 3D sounds; Voice communication; Information tools in car; Visually impaired people

1. Acoustic user interface

1.1. Voice communication system

Recent advances in speech recognition technology have made it possible to enrich human computer interaction (HCI) with voice information. Careful attention must be paid when implementing voice communication in order to create a natural dialogue between user and computer (Fig. 1)

Recently, in Japan, it is not unusual to find a voice input device bundled when purchasing a personal computer (PC).

1.2. 3D sound system

It is true that humans depend on visual sense for 80% of external information, and it is natural that the visual sense is most important during on work with display units. However, making more use of human auditory sense is thought to be worthwhile for several reasons. In particular, the auditory sense can receive audible information from any direction.

This means that information can be displayed by 3D sounds in locations beyond the range of a visual display (Fig. 2).

A 3D sound system can artificially generate an origin for a sound in any direction, in comparison with conventional 2D sound systems which can generate localization only in limited directions (Fig. 3).

The latest advances in computer technology have improved auditory display capability, and real-time sound rendering systems are coming into practical use (Fig. 4). However, despite the advances in technology, little work has been performed developing the human–machine interface (HMI) design of 3D sounds.

1.3. Acoustic user interface

There have been several studies on the spatial interface. One of the studies is 'DynamicSoundscape' at MIT ([1]), which presents a browsing environment that provides the spatial interface for temporal navigation of audio data taking advantage of human abilities of simultaneous listening and memory of spatial location.

In contrast, the author, here, puts forward a HMI that provides a unique concept called acoustic user interface (AUI). The concepts are as follows (Fig. 5):

1. Compared with the conventional audible interface which has monaural or stereo localization, information is displayed effectively by spatially categorized 3D sound localization.
2. Information is categorized not only spatially, but also with other characteristics; sound character (timbre) is used as a metaphor of 'soundscape' in daily living. Specifically the AUI provides features called 'bank counters'.

0141-9382/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0141-9382(02)00011-2

Fig. 1. HUME (NTT Communication Science Research Laboratories, Japan).

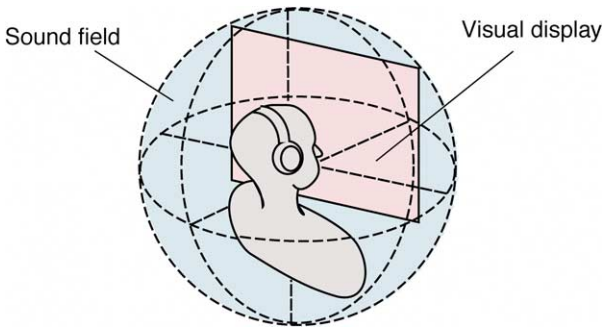

Fig. 2. Comparison of recognition range.

For example, the user can contact one of the virtual personnel who are in charge of a service, such as driving directions.

3. AUI can be used complementarily together with other human interfaces; visual, haptic, etc. For example, haptic devices should be integrated for use by visually impaired people.

When designing a multimodal HMI, the human function allocation must be considered. To reduce the load on the visual sense, the other sensory channels must be utilized. With an AUI, information can be removed from the visual channel and presented through the auditory channel. The head-related transfer function (HRTF) is used to create the impression of sounds arriving from any direction, either inside or outside the visual field of view. As well as reducing the loading on the visual sense, aural images can be projected coincident with the location of images displayed on the visual displays, in order to co-ordinate and communicate more effectively multimedia information.

2. AUI for vehicle information devices

2.1. Why in-vehicle?

When driving a car, or other vehicle, looking at a visual display can distract and cause traffic accidents. However, advances in intelligent transport systems (ITS) are bringing about changes in the interior of the car [2]. And, a lot of information tools which include visual displays are being installed to obtain several information, such as car navigation, traffic information, Email, E-commerce, telecommunication, or so. The worry is that safety is being compromised. It may be possible to reduce the risk of such problems by using AUI and taking advantage of the human allocation function.

2.2. Evaluation of voice communication

Voice communication is one of the characteristics of the AUI. A previous study [3] indicates that by applying conventional voice communication systems, inattentive driving was not effectively reduced. One of the factors of this problem was thought to be that the operator was half-compelled to view the result of speech recognition and

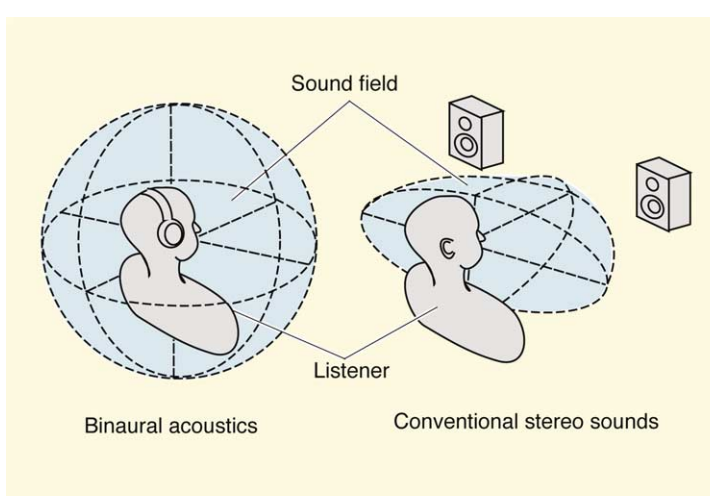

Fig. 3. Concept of 3D sounds.

Fig. 4. Real-time sound rendering system, HURON (Lake Technology Ltd, Picture: Timelord Ltd, Japan).

messages from the system on the visual display. Thus, the authors intended to evaluate the underlying factors.

In this section, the authors report on a qualitative measurement of what kinds of communication error can occur with a voice communication interface [4,5].

Human–machine interface. Fig. 6 shows the HMI between a car-driver and a car navigation system with a voice control function. The elements of each system relate specifically to each other. We can study the usability of a product, to evaluate its HMI.

Evaluation methods: the Story Boarding Evaluation System. The voice communication system in this case was powered by hierarchically structured software. It is important to consider both the flow of the software and the human behavior—especially the errors—corresponding to the flow. The Story Boarding Evaluation System developed by the authors is characterized by multidimensionally evaluating correlation between the flow of the software algorithm and human behavior (Fig. 7).

This evaluation system comprises two main subsystems: the recording system and the evaluation system. The recording system records a multimodal communication across the HMI, using digital images and 3D sounds. The multimedia information, made up of images and 3D sounds, thus recorded are mapped and evaluated by the evaluation system.

In a first pass, images and 3D sounds are edited by the original dual monitoring system. Through this system, both the flow of information on the navigation monitor and the behavior of the subjects are recorded.

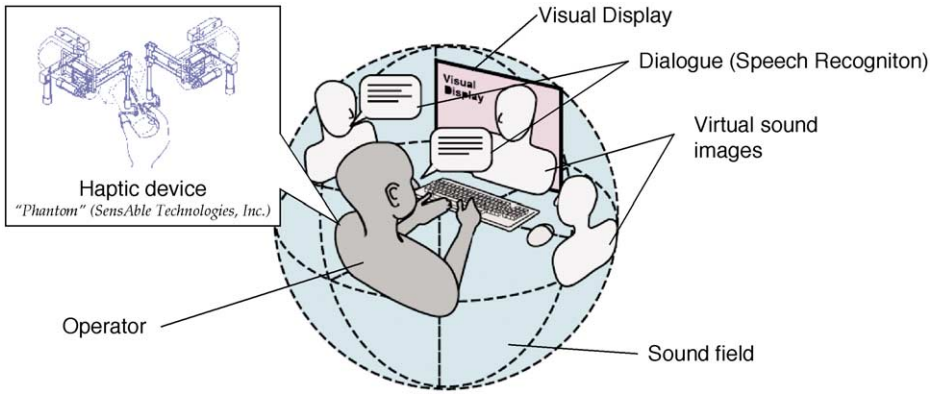

Fig. 5. Concepts of AUI.

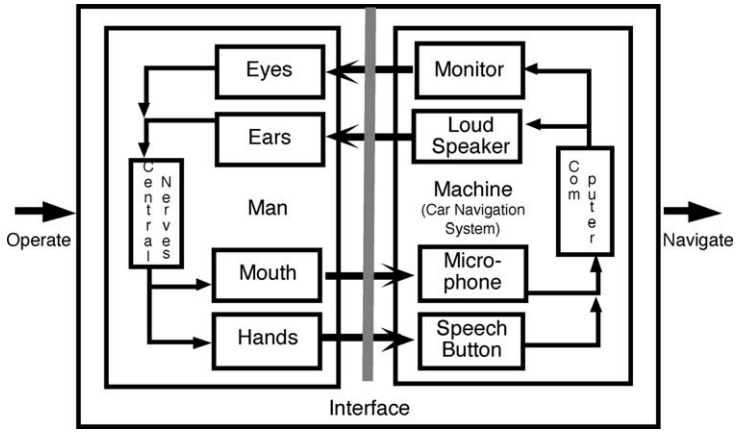

Fig. 6. Human–machine interface.

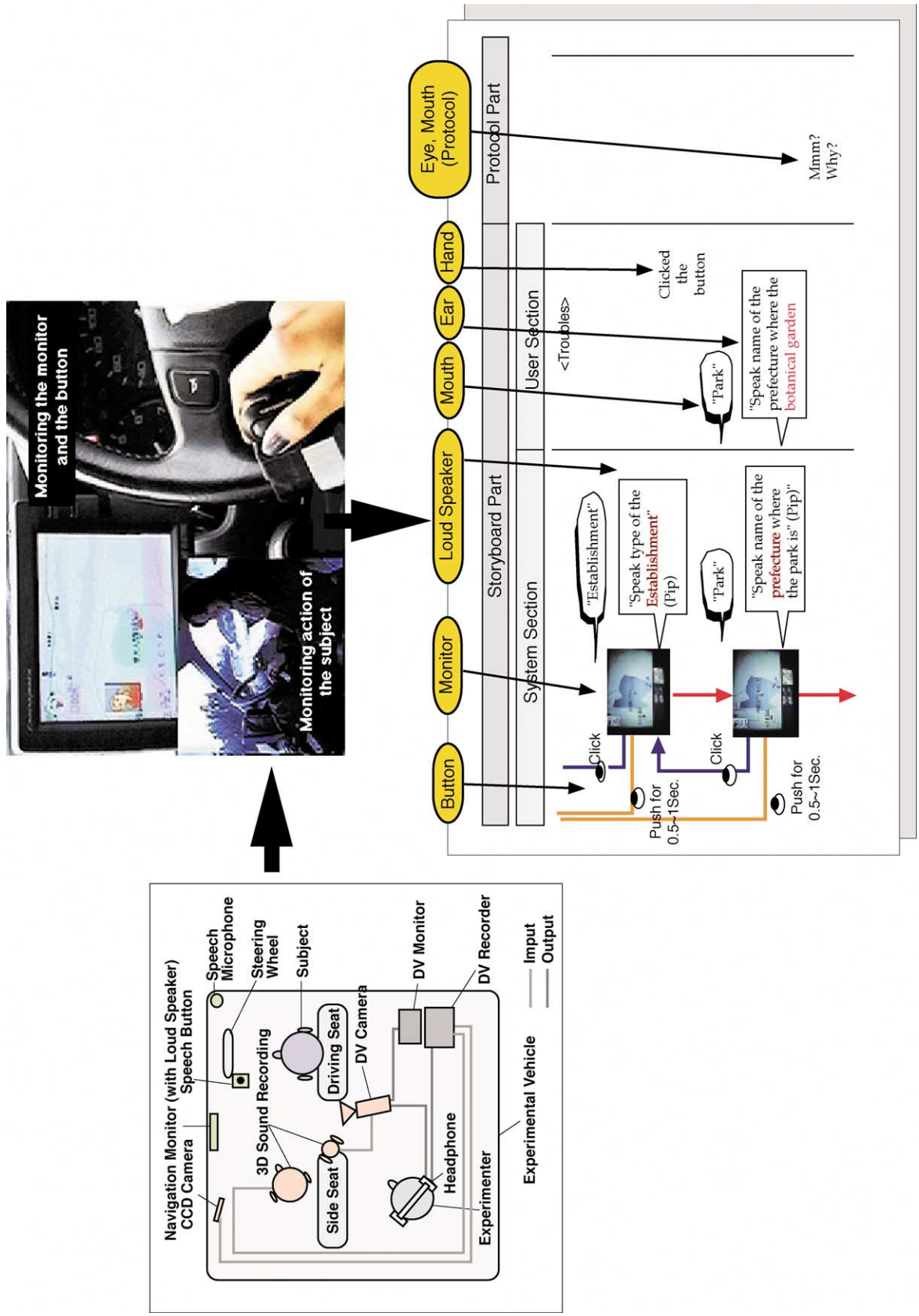

Fig. 7. Story boarding evaluation system.

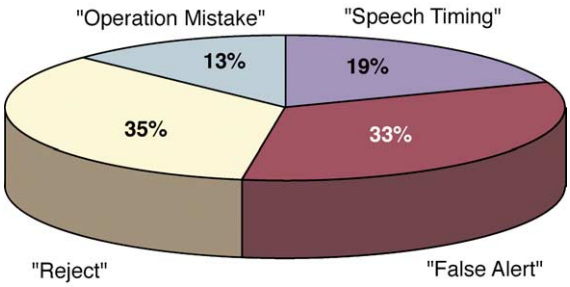

Fig. 8. Error types and percentage.

In a second pass, the results of the observation are mapped on the Story boarding evaluation sheet, consisting of a System Section and a User Section. In the System Section, the flow of information displayed by the navigation system monitor and the loud speakers is described. In the User Section, the subject's behavior—pressing buttons, or spoken responses—are described. Finally behavior and communication error patterns are clustered, and the appearance ratios are calculated.

Evaluation of voice communication. Three car navigation systems were evaluated by the Story Boarding Evaluation System. Four types of error were identified: 'Operation Mistake' error; 'Speech timing' error; 'Reject' error; and 'False Alert' error were clearly isolated (Fig. 8).

Three types of error-loop mechanisms were evaluated in

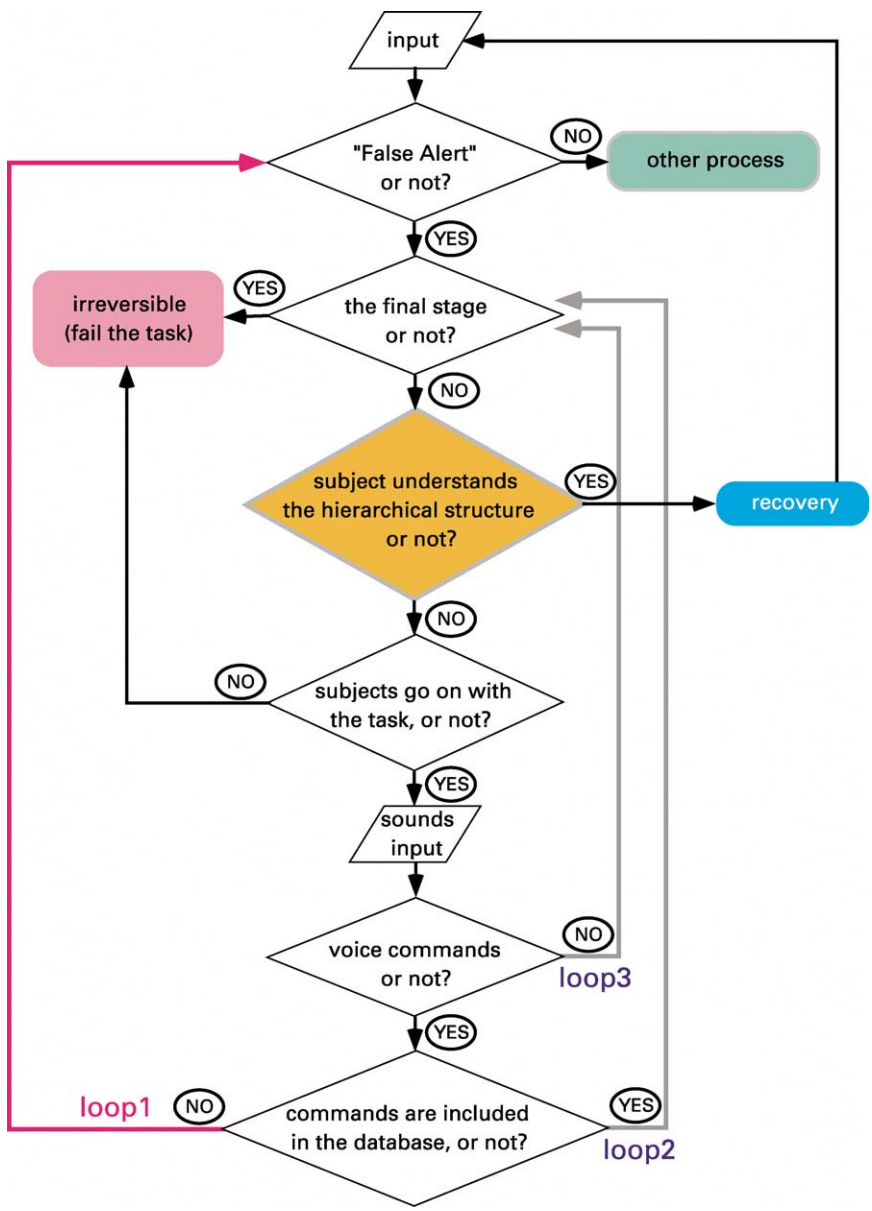

Fig. 9. Flow chart of the error-loop mechanism.

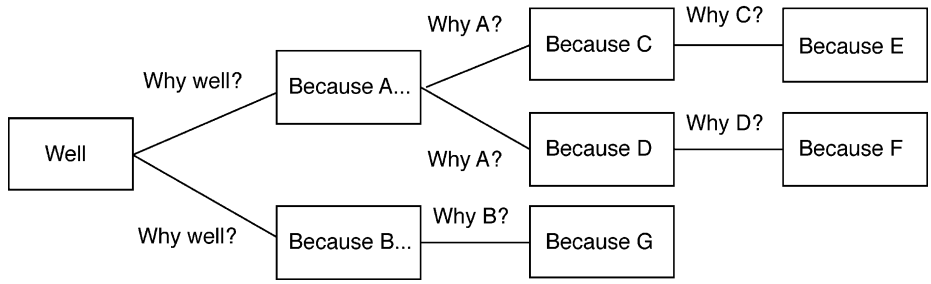

Fig. 10. Personal construct models.

depth. We found that one of the loops, triggered by the false alert, caused a fatal error (Fig. 9).

This serious error-loop suggests that the hierarchical structure of the software interfaces creates the possibility for the user to get lost in the task flow because of poor tabulation of sound information. At this point, natural language processing is thought to be applicable. Alternatively, it is proposed that less complex task flows are used, simplifying the AUI.

2.3. Evaluation of 3D generation of sounds

In addition to voice communication, the 3D presentation of sounds is an important element of the AUI. Most conventional auditory displays are a subsystem of the visual display although some of them contain a little spatial information. However, higher adaptability in sound localization was identified in a spatial 3D sound field, when compared with that in a conventional 2D sound field [4]. In addition, the human auditory sense is able to gather environmental information from any direction instead of gazing at one place.

In order to clarify some basic human factors for HMI design using spatial 3D sound information, the effect on usability of the number of different sound locations and timbres was evaluated [6].

The experimental measurements made were performance on a short-term memory task, a subjective mental workload evaluation and a protocol. For the mental workload, NASA-TLX was applied. To obtain the protocol, the evaluation grid method was applied.

National aeronautics and space administration task load index (NASA-TLX). NASA-TLX is a subjective workload assessment tool originally developed by NASA [8]. NASA TLX allows users to perform subjective workload assessments on operators working with various human-machine systems. NASA-TLX is a multi-dimensional rating procedure that derives an overall workload score based on a weighted average of ratings on six subscales. These subscales are Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort, and Frustration. It can be used to assess workload in various human–machine environments such as aircraft cockpits; command, control, and communication (C3) workstations; supervisory and process control environments; simulations, and laboratory tests.

Table 1
Experimental design

Direction	Timbre		
		1	3
	1	A Group	B Group
	3	C Group	D Group
	5	E Group	F Group

The evaluation grid method. The evaluation grid method was originally developed for architecture and has recently been applied to ergonomics [7]. In the case of a radio channel changer, for example, such paired adjectives as the button size being 'big–small' or the feeling the operation to be 'smooth–clumsy' can be connected to form an image like 'this channel changer has smooth operating feelings because the buttons are big'. This model of human cognition is called the construct system.

The evaluation grid method is one of interviewing to obtain the personal construct systems (Fig. 10). This method is intended to obtain the mechanisms by which a human recognizes and structures understanding of the environment or human relations.

Evaluation. The experiment was designed with several observations per combination in order to be able to apply the pair comparison method at the same time as the evaluation grid method. The factors, the levels and the experimental conditions are shown in Table 1.

For the single location condition, the speech was binaurally recorded with a dummy head microphone from a source at the same height as, directly in front of and 60 cm distant from the microphone. To obtain the five-locations condition the speech was recorded as already explained and additionally from locations at the same height and distance (60 cm), but offset by 60° to the right and left, respectively.

To obtain the single timbre condition, the voice of an adult male was used. For the three-timbres condition, voices of three adult males were used.

The arrangement of the locations and timbres are shown in Fig. 11. Displayed stimuli consisted of the names of the objects belonging to one of three categories; animals,

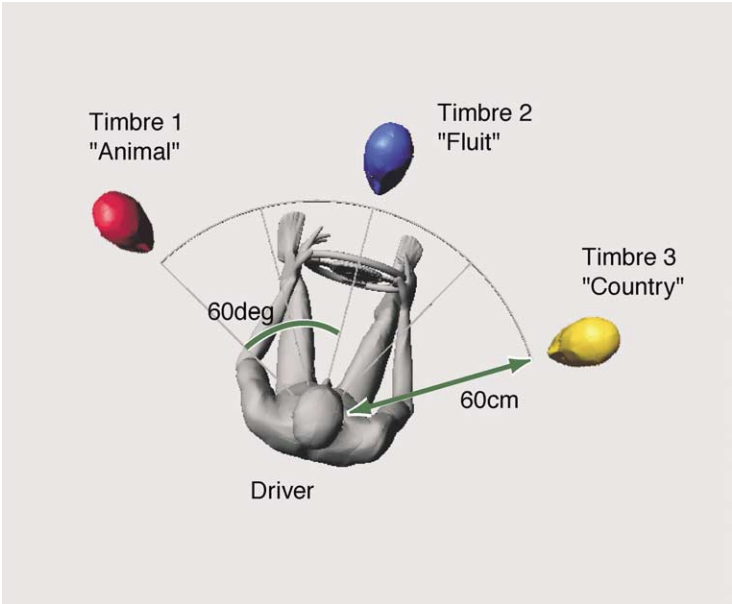

Fig. 11. Arrangement of the source and timbre.

vegetables and countries. For example, under the condition of the D group, each category's contents were displayed randomly by three types of timbre and three-directions. The same voice (timbre) was used for all names in the same category.

The arrangement of single sound location condition (A and B group) is very similar to that of conventional monaural sound display. In general, the loud speaker of the car navigation systems is installed in the visual monitor, and in many cases, monaural sound (single direction) is displayed.

The experimental tasks were part of dual-task performance to recreate the characteristics of information cognition in a car-driving environment. The sub-task was the tracking, with a mouse and cursor, of a car driving ahead on the video screen of a driving simulator (Fig. 12).

The main task was the short-term memory test. The test required the subjects to memorize as many words as possible from 20 words displayed randomly at 1 s intervals. A total of four trials were presented at random for each test. The subjective measurements were recorded after each trial.

Results of the short-term memory test. A two factorial ANOVA was applied to the scores of the short-term

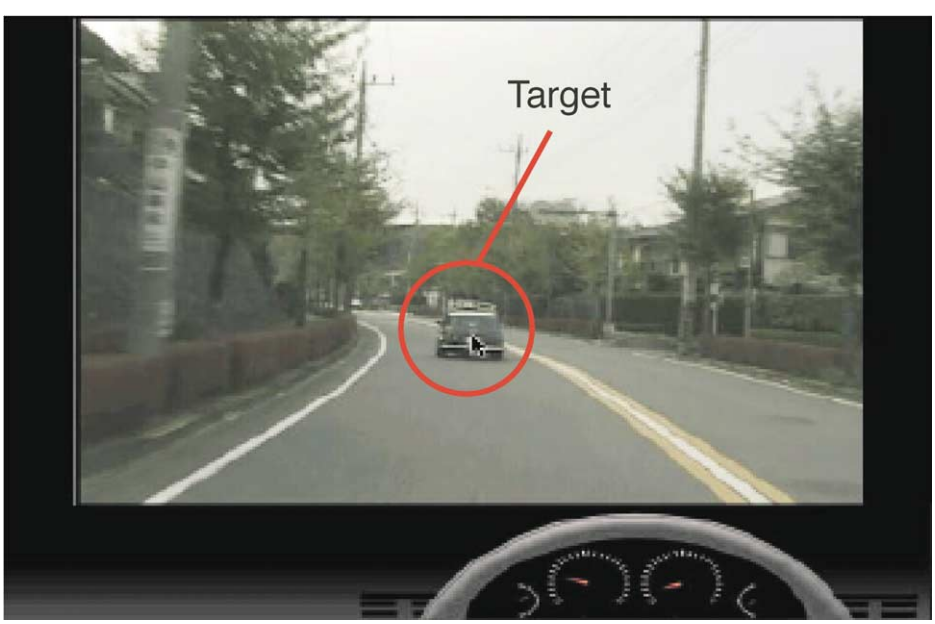

Fig. 12. Driving simulator.

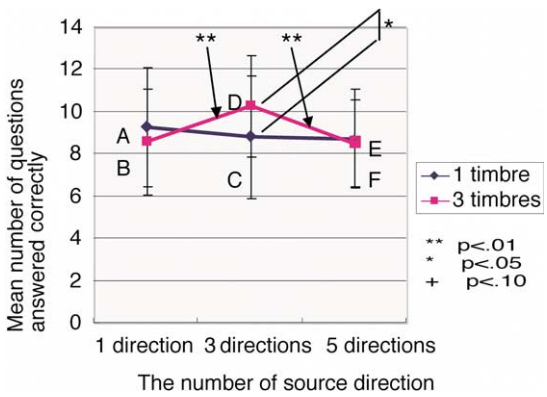

Fig. 13. Chart of the interaction (short-term memory test).

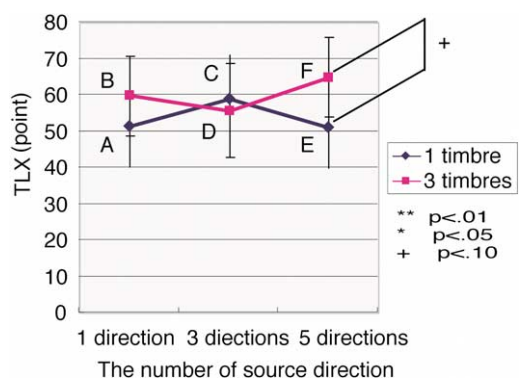

Fig. 14. Chart of the interaction (mental workload).

memory test (Fig. 13). As a result, the significant interaction was found between each factor, $F(2,24) = 3.12$ $p < 0.01$. As the result of the analysis of interaction, the simple main effect of the two-timbres level within the number of direction factor was significant, $F(2,24) = 4.73$ $p < 0.01$. The simple main effect of the number of timbre factor at the three-directions condition was significant, $F(2,24) = 5.27$ $p < 0.05$.

A multiple comparison by LSD shows that the significant difference of the number of the direction factor at the three-timbres level was found.

At the same time, a significant difference was found between the number of three-directions and the five-directions level at the two-timbres condition.

Thus, the three-directions condition was associated with a significant improvement in performance of short-term memory. In particular, the condition of three-timbres and three-directions resulted in the highest performance of short-term memory.

Results of mental work load. A two factorial ANOVA was applied to the scores of the mental workload evaluation. The interaction was found between each factor; the multiple comparisons were applied (Fig. 14). The three-timbre-levels at five directions condition, was found to provoke a higher workload evaluation than the single timbre condition at five directions. No significant difference was found among other conditions.

Results using the evaluation grid method. This method found that with the increase in spatial information the subjects felt it was easier to memorize. On the other hand, less spatial information was felt to make it easier to listen. One of the construct models is shown in Fig. 15. The pair comparison method was also applied. As a result, the order of preference of the groups was as D > A > B > C > F > E. This result is thought to reinforce the results of the objective evaluation above.

Conclusion. In this study, the authors evaluated an AUI applied to car-information tools in order to discover the human factors important for the design of such an AUI. First, the authors report on a qualitative measurement of the types of communication error that can occur with a voice communication interface. Four types of communication errors and three typical types of error-loop mechanisms triggered by false alerts were evaluated.

Secondly, the authors discovered some basic human factors of HMI design using spatial 3D sound information. The effect of the number of sound locations and timbres on usability was evaluated. It was found that for recognizing audible information while driving a car, to display the sound information from several different directions contributed positively to the cognition of sound information and lowered mental workload. This result indicates that displaying spatial 3D sound information is more helpful than the conventional monaural sound information in designing the HMI.

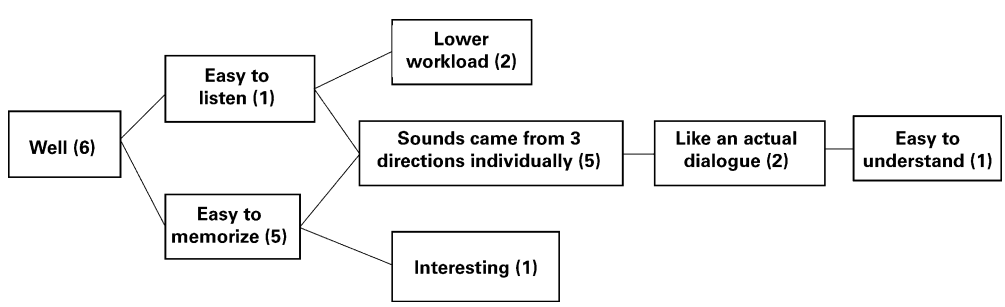

Fig. 15. A construct model (D > A).

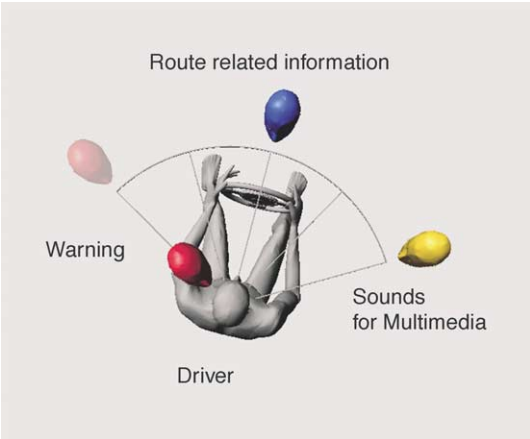

Fig. 16. The priority allocation system by AUI.

3. Future prospects

The authors have been developing an advanced design of AUI.

For example, a priority allocation system which is displayed by spatially categorized 3D sounds, powered by an AUI, is under development (Fig. 16).

In the medium- and long-term, continual evaluation and consideration of concepts will be carried out for the overall development of the AUI.

In the short-term, the authors will apply the AUI to several fields other than ITS. For example, AUI for public services will be investigated.

AUI offers the possibility to improve accessibility and reduce barriers to information. Conventional visual display terminals with graphical user interfaces offer poor accessibility for visually impaired people. Including AUI in the interface design would improve the accessibility of the HCI for visually impaired users.

In addition, the use of voice communication would reduce the importance of using pointing devices, such as the mouse, and achieve hands-free interaction. In such a way, people who have reduced performance in the upper limbs could enjoy surfing the Internet through an AUI.

Acknowledgements

We would like to thank Ms Y. Hattori, Mr T. Nakahara and Mr F. Suzuki for grateful assistance. We wish to acknowledge valuable discussions with Prof. T. Yasuzuka and Dr H. Nishimura (Kanagawa Institute of Technology).

References

[1] M. Kobayashi, C. Schmandt, Mapping time to space for audio browsing, Transactions of Information Processing Society of Japan 39-5 (1998).

[2] Japan Ministry of Land, Infrastructure and Transport Road Bureau ITS Homepage, http://www.mlit.go.jp/road/ITS/index.html, 2001.

[3] J. O-sugi, H. Takao, K. Sakai, K. Noro, The characteristics of eye movement in the voice control car navigation system, Proceedings of the 29th Annual Meeting of the Kanto-branch, Japan Ergonomics Research Society, 1999, pp. 40–41.

[4] H. Takao, K. Noro, A study on learning of direction of sound images in virtual reality, Transactions of the Technical Committee on Psychological and Physiological Acoustics, Acoustical Society of Japan H99-H1 (1999).

[5] K. Sakai, H. Takao, J. O-sugi, K. Noro, Analyzing troubles on voice control car navigation system by protocol method, The Japanese Journal of Ergonomics 35 (Suppl) (1999) 252–253.

[6] H. Takao, J. O-sugi, H. Ishii, K. Noro, Utilizing acoustic information for the information tools on a car, Ergonomics of Human-System Interaction (2001).

[7] J. Sanui, Visualization of users' requirements: introduction of the evaluation grid method, Proceedings of the Third Design and Decision Support Systems in Architecture and Urban Planning Conference, vol. 1, 1996, pp. 365–374.

[8] The Human Systems Information Analysis Center (HSIAC), NASA Task Load Index (TLX), http://iac.dtic.mil/hsiac/products/tlx/tlx.html, 2001.