Differences in Mental Workload when Processing Visual and Auditory Information

—Effect of Presentation Method, Recall Type, and Time Period for Analysis—

Kazumoto MORITA*

The author conducted a laboratory experiment to estimate the mental workload of subjects as they acquired visual and auditory information. The following factors were examined using a within-subjects design experiment.

- Presentation method: The 3 modes of presenting both auditory and visual information simultaneously, presenting only auditory information, and presenting only visual information;
- Whether information was recalled vocally or silently after it was presented; and
- Three time periods for analysis: before, during, and after presentation.

As a subsidiary task, subjects were made to rotate a mounted disk 6 cm in diameter one full turn every 2 s. The variance in time to rotate the disk was examined to clarify the differences in mental workload between experimental conditions. The results were as follows. No difference in mental workload was found among the 3 presentation modes. Moreover, it was clearly shown that mental workload was highest when subjects replied vocally.

Key Words: mental workload, visual information, auditory information, automobile, dual task method

1. Introduction

Information devices such as navigation systems in automobiles are spreading rapidly, increasing the opportunities for drivers to acquire information while driving. These navigation systems also have increasingly higher-level functions; in addition to giving road guidance they are expected in the future to provide a variety of information by serving as an Internet terminal. While such provision of various information increases drivers' convenience and comfort, there is also concern that the actions performed to acquire information may have a deleterious effect on driving operation.

The author, therefore, carried out a basic indoor experiment to estimate the level of the mental workload when information was transmitted to the driver from a device, using a dual task method.

Here, not only visual methods but also auditory ones using audio devices can be employed for transmitting information to the drivers¹⁾. A characteristic of presenting auditory information is that it does not induce the driver to take his eyes off the road, although some time is required

until all the information is transmitted. In presenting information, therefore, it is necessary to take advantage of the respective characteristics of visual and auditory information²⁾.

Considering the method of presenting information, it would seem that the mental workloads placed on the driver may differ depending on whether the information is presented auditorily or visually. To clarify this issue, the author examined the differences in mental workload depending on the presentation method.

In experiments on memory and recall related to mental workload, subjects are commonly made to vocally recite memorized information. However, when driving an automobile, drivers do not need to recite each and every item after they have memorized necessary information. Therefore, the level of the mental workload on subjects was also investigated when they recall memories vocally. The findings for mental workload when speaking may also be applicable to investigations of safety when a driver is, for example, using a mobile phone while driving.

2. Experiment

2.1 Outline of experiment

In the present laboratory experiment, subjects were presented visual and auditory information in the following 3 ways.

* National Traffic Safety and Environment Laboratory,
7-42-27 Jindaijihigashimachi, Chofu
(Received December 27, 2004)
(Revised July 11, 2005)

- (1) Presentation of both visual and auditory information (dual presentation).
- (2) Presentation of auditory information only (auditory presentation).
- (3) Presentation of visual information only (visual presentation).

The presented information was place name groups consisting of 6 place names each (described below). The subjects rested their chins in a chin rest to fix head position, and memorized the presented place names while viewing a computer display at a distance of approximately 0.6 m, or listening to a computer speaker (Photo. 1). Both visual and auditory information was presented for a duration of 6 sec, and the subjects were asked to accurately recall as many of the place names as they could. In addition, the effects of utterances were examined using 2 methods: having the subjects recall the memorized place names vocally, and having them recall the place names silently in their heads.

In this study a dual task technique was used to estimate the level of the mental workload of subjects. Various subsidiary tasks can be devised^{(3)~(6)}, but in the present experiment subjects were instructed to rotate a disk 6 cm in diameter (Photo. 2). The time to rotate the disk was recorded by a computer and used as the evaluation index. The rotation time was not measured only during the period of information presentation, but also for 6-sec periods before and after the presentation of information.

Photo. 1 A subject watching display screen

Photo. 2 Subsidiary task: rotating a disk 6 cm in diameter

This experiment was based on a within-subjects design for the following 3 factors:

- (1) Presentation method (dual presentation, auditory presentation, visual presentation)
- (2) Whether or not subjects responded vocally following the presentation
- (3) Analysis period of rotation time (before presentation, during presentation, after presentation)

2.2 Information presented

The information presented to subjects was composed of 30 different groups of place names (6 names per group) from Tokyo and the surrounding regions.

For the visual presentations, a display area 6 inches in size was established on the computer screen, in which black characters were displayed against a white background in 2 rows with 3 names each (Photo. 3). A buzzer sounded for 0.3 sec starting 0.8 sec before the information appeared. Then, after a 0.5 sec interval the place name group was shown for 6 sec. After the presentation ended there was another 0.5 sec interval followed by the buzzer sounding again for 0.3 sec. The subjects were then asked to recall the place names. A time breakdown of the entire presentation flow is shown in Fig. 1. The time at the start of the presentation was recorded by the computer, as well as the time 6.8 sec before and 6.8 sec after the start of the presentation.

For the auditory information, 6 place names that had

Photo. 3 Example of displayed image of place names

Fig. 1 Time sequence for presentation

been recorded in advance on the computer were played back in order. The overall time composition was the same as in Fig. 1, with the buzzer sounding before and after the presentation.

The dual presentation was made using the same place name group both visually and auditorily, with the displayed place names read out in order from the upper left to the right.

2.3 Subsidiary task

As a subsidiary task subjects were instructed to rotate a disk at a constant speed and rate of 1 rotation every 2 sec. They first practiced to become accustomed to the feeling of rotating the disk one rotation every 2 sec. During this proficiency practice the time for one rotation was shown on the computer screen, so that by watching the time on the computer they could acquire the sense of rotating the disk at the speed of once per 2 sec. This condition of 1 rotation/2 sec was established in a preliminary experiment from the 2 perspectives of rotating the disk slowly enough so that variance in the rotation speed was seen but also rapidly enough so that as much data as possible could be obtained within a limited measurement time.

A pulse signal was sent from the rotating disk to the computer every half turn, and time differences between the pulse signals (half rotation times) were obtained and used in the post-experiment analysis. Since the time was 1 rotation every 2 sec, it was expected that the half rotation time would be 1 sec.

The author also made measurements in 2 trials for each subject under a control condition, in which the subject rotated the disk only, without presenting place name groups. The later analysis was conducted based on the half rotation times measured in these trials. In a single trial under the control condition, 60 half rotation times (about 1 minute) were obtained, so 120 individual data were analyzed for each subject.

2.4 Experimental methods

The experiment was conducted with paired vocal and silent recall. The following is an example of the experimental procedure.

- (1) Control condition
- (2) Dual presentation (vocal recall)
- (2') Dual presentation (silent recall)
- (3) Auditory presentation (vocal recall)
- (3') Auditory presentation (silent recall)
- (4) Visual presentation (vocal recall)
- (4') Visual presentation (silent recall)
- (5) Control condition

The control conditions were used at the beginning and

end of the experiment.

During the experiment, items (2), (3), and (4) above were set in random order for each subject to prevent presentation order bias. In addition, the presentation method included the 2 modes of vocal and silent recall. Half of the subjects underwent the vocal recall trials first, and the other half the silent recall trials first. Prior to the experiment trials with each presentation method, in the case of both vocal recall and silent recall, subjects went through 2 practice trials using practice place name groups.

Each subject underwent 5 trials with each presentation method with vocal and silent recall, respectively. The total number of trials for each subject was therefore 30 (3 presentation methods $\times$ 2 response modes (vocal and silent) $\times$ 5 repetitions). These 30 trials corresponded to the 30 different place name groups mentioned above.

Since there was concern that when subjects were not giving a vocal response they would make less effort to memorize the place name groups presented, they were told beforehand that they may be asked to respond at any time. After the post-presentation 6-sec period for measurements of half rotation time had passed, random requests for responses were made 1 time for each of the 3 presentation methods. In these cases the responses were given about 6 sec later than with the normal vocal recall.

It was also predicted that the subjects' sense of making 1 rotation of the disk every 2 sec would change during the time of the experiment, so before trials with each of the presentation methods the subjects were asked to practice to re-familiarize themselves with the sense of making 1 rotation every 2 sec.

The twenty subjects (10 males, 10 females, mean age 32.9 years and S.D. 4.9) were recruited from outside our institution and paid for their participation. The time each subject spent at the experiment site was about 1 h in total, including time for explanation, familiarization practice, and trials under the control condition.

3. Results

3.1 Error analysis

When a subject responded with a place name that was not actually presented, it was considered to be an error. Under the vocal recall conditions, subjects were asked to respond with a total of 600 individual place names under each presentation condition (6 place names in a single presentation $\times$ 5 repetitions $\times$ 20 subjects = 600). In contrast, under the silent recall condition a response was requested only once out of the 5 repetitions. Therefore, in this case there were possible responses with 120 place

names (6 place names in a single presentation $\times$ 20 subjects = 120). As shown in **Table 1**, subjects memorized and responded accurately most of the time, with few mistakes under any of the 3 conditions. From now on, correct responses will be analyzed.

3.2 Number of place names

The mean number of place names given by vocal recall following presentation is shown in **Fig. 2** with the solid circles for each of the 3 presentation methods. Subjects were able to vocally recall about 4 place names with all 3 presentation methods. A *t*-test was conducted for the difference in mean number of place names with each of the presentation methods ($p < 0.05$), and a significant difference was found between dual and visual presentation. There were no other significant differences between any other conditions.

Under the silent recall condition, subjects were asked to silently recall the place names in their heads after presentation. For each presentation method, they were asked randomly to respond vocally to one of the presentations. The mean numbers of place names given at these times are shown with the open circles in **Fig. 2**. The number of responses (about 3.5) was smaller than that when the subjects were asked to respond vocally. A *t*-test ($p < 0.05$) to determine whether there was a significant difference between the vocal recall and silent recall condi-

tions revealed no significant difference with visual presentations, but significant differences with dual or auditory presentations. The reasons fewer responses were given in the silent recall condition than in the vocal recall condition may be that the subjects' memory dissipated in the 6 sec that had passed from the end of the presentation⁷⁾.

3.3 Half rotation time under the control condition

The mean value for all half rotation data under the control condition was 1.03 sec (S.D. 0.20). This is very close to 1 sec, and corresponds with the experimental instructions. However, the mean and standard deviation for half rotation time for individual subject revealed some differences among subjects, as shown in **Fig. 3**.

3.4 Results of half rotation time analysis

Data for the half rotation time were analyzed for the 6 sec of the presentation, 6 sec before the presentation (precisely, before the buzzer started), and 6 sec after the presentation (precisely, after the buzzer stopped). Of the 6 sec, the first and last data measured were excluded so that only valid data were used.

Here, the conditions for each analysis period before, during, and after presentation with each presentation method are termed the trial conditions. Thus, for each subject there were 18 trial conditions (3 presentation methods $\times$ 2 modes of vocal and silent recall $\times$ 3 analysis periods). Each subject performed 5 repetitions under each trial condition, and the data were analyzed based on the results of those 5 repetitions.

Since subjects did not necessarily rotate the disk properly at a uniform speed, the amount of valid data for each subject under each trial condition varied. Considering simplicity of analysis, the same number of data was extracted under each trial condition for use in the analysis. Thus, since the minimum number of data through all trial conditions was 15, this same number of 15 data was extracted randomly for each subject under each of the trial condition.

Fig. 4 shows the mean and standard deviation for half rotation time with each of the presentation methods. The

Table 1 Numbers and percentages of errors

Presentation method	Number of errors (A) and percentage (A/total no. of presentations 600) when subjects recalled vocally	Number of errors (B) and percentage (B/total no. of presentations 120) when subjects recalled silently
Dual presentation	11 (1.8%)	3 (2.5%)
Auditory presentation	7 (1.2%)	2 (1.7%)
Visual presentation	20 (3.3%)	5 (4.2%)

Fig. 2 Number of place names answered according to presentation method

Fig. 3 Half rotation time under control condition

results are shown for vocal recall (solid circles) and silent recall (open circles), respectively. Overall, the values were smaller than the half rotation time values under the control condition. Thus, the subjects tended to rotate the disk faster than when under the control condition in which there was no mental workload.

As seen in Fig. 4, although a difference was found between the vocal and silent recall conditions after the presentation, the difference between the trial conditions was small and no clear trend was seen. The reason for this is thought to be that mean values were taken for half rotation time. Observing the subjects during the experiment, it was found that under a mental workload they tended to rotate the disk at an uneven speed rather than smoothly. Therefore, it would seem that comparisons of the variance in half rotation time are more appropriate than comparisons of mean time in investigating the effects of mental workload.

Thus the variance in half rotation time (rotating time variance) was obtained based on the mean squared deviation using the following equation. The reference value was the mean half rotation time under the control condition for each subject.

$$y_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (x_{ij} - a_i)^2}$$

Here,

- y_i : Rotating time variance for each subject
- x_{ij} : Half rotation time for each subject under each trial condition
- a_i : Half rotation time for each subject under the control condition
- N : Sample size for each trial condition (in this study, 15)

Fig. 5 shows the mean rotating time variance for each of the 3 presentation methods. Vocal recall (solid circles)

and silent recall (open circles) responses are shown. Regardless of presentation method, the rotating time variance before presentation was small. After the presentation, the rotating time variance was larger with vocal recall than with silent recall.

4. Discussion

4.1 Results of analysis of variance

Using a within-subjects design, an analysis of variance was conducted with rotating time variance as the dependent variable, and the 3 factors of A: presentation method (A1: dual presentation, A2: auditory presentation, A3: visual presentation); B: recall type (vocal (B1) or silent (B2)); and C: time period for analysis (C1: before presentation, C2: during presentation, C3: after presentation) as independent variables. The results showed a main effect for the factors of vocal or silent recall ($F(1, 19) = 10.98, p < 0.01$), and time period for analysis ($F(2, 38) = 19.58, p < 0.01$), respectively. Significant interaction was seen between vocal or silent recall and time period for analysis ($F(2, 38) = 9.85, p < 0.01$).

An analysis of simple main effect for this interaction (Table 2) revealed a significant difference ($p < 0.01$) for

Table 2 Results of analysis of simple main effect

S.V.	SS	df	MS	F
B at C1	0.0006	1	0.0006	0.42
S×B at C1	0.0292	19	0.0015	(n.s.)
B at C2	0.0064	1	0.0064	2.25
S×B at C2	0.0541	19	0.0028	(n.s.)
B at C3	0.1170	1	0.1170	15.59
S×B at C3	0.1427	19	0.0075	(p<0.01)
C at B1	0.2367	2	0.1183	21.67
S×C at B1	0.2075	38	0.0054	(p<0.01)
C at B2	0.0429	2	0.0214	6.87
S×C at B2	0.1186	38	0.0031	(p<0.01)

Fig. 4 Half rotation time according to three presentation methods

Fig. 5 Rotating time variance with respect to three presentation methods

Table 3 Results of multiple comparison tests

C at B1 (MSe=0.0054, $p<0.05$) (LSD=0.0275)	C1<C2 ($p<0.05$) C1<C3 ($p<0.05$) C2<C3 ($p<0.05$)
C at B2 (MSe=0.0031, $p<0.05$) (LSD=0.0208)	C1<C2 ($p<0.05$) C1<C3 ($p<0.05$) C2=C3 (n.s.)

Fig. 6 Results of analysis of variance with rotating time variances

factor B at C3 (after presentation). However, there was no significant difference in factor B at C1 (before presentation) or C2 (during presentation).

In addition, a multiple comparison test was conducted using the least significant difference (LSD) method for factor C in mode B1 (vocal recall) (Table 3). The results revealed significant differences between all analysis periods. As shown in Table 3, a multiple comparison test for factor C in mode B2 (silent recall) revealed significant differences between C1 (before presentation) and C2 (during presentation), and between C1 (before presentation) and C3 (after presentation), but not between C2 (during presentation) and C3 (after presentation). Fig. 6 is a graphic representation of these results.

The results of these tests revealed the following.

- (1) Factor A was not shown to exert an effect on rotating time variance. Thus, no difference in mental workload was found according to the 3 presentation methods. Before presentation, since subjects were simply waiting, no difference owing to the presentation method would be expected. However, even during presentation there was no apparent difference in mental workload when subjects were acquiring information visually and auditorily.
- (2) Analysis results for the interaction between factors B and C showed that mental workload was largest when subjects gave a vocal response after presentation. The load was larger when subjects were vocally responding than when they were

Fig. 7 Number of cases where significant differences in ANOVA were recognized out of 20 samplings. Here, A: presentation method, B: recall type, C: time period for analysis (C1: before, C2: during, C3: after)

acquiring the information during presentation. When they were simply recalling the information silently to themselves there was no difference in load compared with that during presentation.

- (3) Given that there was an effect on factor B only at C3 (after presentation), vocal or silent recall was related solely to recall response after presentation; before or during the presentation there was no mental workload difference depending on vocal or silent response. These facts agree with our common sense and are evidence for the validity of the estimation method in this study.

4.2 Number of subjects

The present experimental results are predicted to be related to the number of subjects, and the author shall examine this issue further.

Data for N subjects were randomly extracted from the 20 subjects in this study, and an analysis of variance was repeated for rotating time variance for this group. Here, the N subjects were varied from 5 to 19 subjects, and the number of samplings was 20 times for each number of subjects. An analysis was conducted as to whether or not there was a significant effect of utterances in each of the time periods for analysis. The author examined the cases as follows: interaction was found between factor B (type of recall) and factor C (time period for analysis), and the results of this lower ranked test gave the number of cases in which there was a significant difference for B at C1, C2, and C3. Here, differences were judged to be significant at a significance level of 5%.

The vertical axis in Fig. 7 shows the number of cases in which a significant difference due to recall type was found among the 20 samplings. Before presentation (B at C1) and during presentation (B at C2), there were few instances of significant difference regardless of the number of subjects. As for the effect of utterances after presenta-

tion (B at C3), significant differences were seen in all 20 samplings when the number of subjects was 14 or greater. In the design of the present experiment, about 14 subjects were required to clearly identify significant differences.

For reference, the results of an analysis of main effect for factor A (presentation method) are also shown in the figure. Here again significant differences were seen in few cases, so the same result was obtained regardless of the number of subjects.

5. Conclusion

Differences in mental workload when processing visual and auditory information were examined using a within-subjects experimental design for 3 factors. In this experiment the subjects were made to perform a subsidiary task of rotating a disk, and the variance in rotation time was investigated for the following factors:

- (1) Presentation method; namely the 3 methods of dual presentation of visual and auditory information, auditory information only, and visual information only;
- (2) Vocal or silent recall of information after presentation; and
- (3) Period of analysis: before, during, or after the presentation.

The presented information was place name groups consisting of 6 place names, and the subjects were asked to memorize and recall these names presented for 6 sec.

The main findings were as follows.

- (1) About 4 place names could be recalled after presentation, and there was no marked difference in this number depending on the method of presentation.
- (2) No significant difference was found in mental workload depending on the presentation method when acquiring information during presentation.
- (3) When subjects were made to give a vocal response after presentation, the mental workload was larger than when they recalled the information silently. In addition, the mental workload was also larger when recalling vocally than when

acquiring the information during presentation. Therefore, whether or not subjects are required to give a vocal response is predicted to strongly affect the level of the mental workload.

References

- 1) K. Kishi, K. Kimura and S. Sugiura: Human Factors Considerations for Voice Route Guidance, Proceedings of Annual Conference of the Society of Automotive Engineers of Japan, 936, 93/96 (1993) (in Japanese)
- 2) K. Morita, M. Sekine, T. Okada and J. Mashiko: Amount of Auditory Information Which Driver Can Obtain at a Time While Driving, Proceedings of Annual Conference of the Society of Automotive Engineers of Japan, No. 5-4, 17/22 (2004) (in Japanese)
- 3) J. Walker, E. Alicandri, C. Sedney and K. Roberts: In-Vehicle Navigation Devices: Effects on the Safety of Driver Performance, Proceedings of the Vehicle Navigation and Information Systems Conference, Part 1, 499/525 (1991)
- 4) L. Harms and C. Patten: Peripheral detection as a measure of driver distraction. A study of memory-based versus system-based navigation in a built-up area, Transportation Research, Part F6, 23/36 (2003)
- 5) L. Tijerina, S. M. Kiger, T. H. Rockwell and C. Tornow: Workload Assessment of In-Cab Text Message System and Cellular Phone Use by Heavy Vehicle Drivers on the Road, Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting, 1117/2000 (1995)
- 6) Y. Uchiyama, T. Wakita, S. Kojima and T. Hongo: Measurement of Drivers' Cognitive Workload Using Dual Task Methods, Proceedings of Annual Conference of the Society of Automotive Engineers of Japan, No. 55-02, 5/8 (2002) (in Japanese)
- 7) G. R. Loftus and E. F. Loftus: Human Memory, The Processing of Information, 39/43, Lawrence Erlbaum Associates (1976)

Kazumoto MORITA (Member)

Dr. Morita was born in 1951. He graduated from the graduate school of Kyoto University in 1976 (M. Eng.) and started working for Traffic Safety and Nuisance Research Institute of the Ministry of Transport of the Japanese government. He obtained degree of Dr. in Engineering in 1998. He now is a Principal Researcher at the National Traffic Safety and Environment Laboratory. His research interests include human-machine interface issues between drivers and vehicles.