■ 論 文 ■

Comparative Study of Subjective Mental Workload Assessment Techniques for the Evaluation of ITS-oriented Human-Machine Interface Systems

지능형 교통체계 기반 인간-기계 인터페이스 시스템 평가를 위한 정신적부하 측정방법의 비교 연구

CHA, Doo Won
(Ph. D. Candidate, Industrial Engineering Department, Ajou University)

PARK, Peom
(Associate Professor, Industrial Engineering Department. Ajou University)

목 차

Key Words : ITS, Subjective Mental Workload, NASA-TLX, SWAT, MCH, RNASA-TLX

요 약

교통시스템 및 차량의 인간-기계 상호작용 및 인간공학적 평가에 널리 사용되는 운전자의 시각행위분석, 생체신호측정(심박수, 호흡수, 뇌파, 피부전기반사 등), 부가작업의 수행도분석, 반응시간 분석, 차량조작변수 분석(속도, 핸들조작, 가속페달조작) 등의 방법들과 함께 운전자 판단에 의한 주관적평가(Subjective Mental Workload Assessment)는 기존 시스템의 평가뿐만 아니라, ITS(Intelligent Transport Systems) 기반 인간-기계 시스템(항법장치, Head-Up Display, AHS Flat Panel Display, Changeable Message Sign, Cellular Phone, 등)의 인간-기계 상호작용 평가, 안전도 평가, 경쟁시스템 및 디자인 대안의 의사결정, 설계가이드라인 제작 등을 위한 하나의 표준화된 평가도구로 자리 잡고 있다.

이에 본 연구에서는 가장 많은 적용빈도와 효과 및 감도를 제시하는 주관적 평가방법인 NASA-TLX (National Aeronautics and Space Administration-Task Load Index), SWAT(Subjective Workload Assessment Technique), MCH(Modified Cooper-Harper) scale, RNASA-TLX(Revision of NASA-TLX)를 대상으로 항법장치를 사용한 운전자 부하 비교평가 실험결과를 이용하여, 제시된 방법들의 감도 및 피실험자의 주관적 판단을 통한 상호 비교분석을 실시하였으며, 결과로 위에 제시된 방법들의 사용가이드라인과 ITS 기반 인간-기계 시스템의 인간공학적 평가를 위한 주관적 평가방법의 비교평가결과를 제시하였다.

Ⅰ. Introduction

Recently, several types of ITS-oriented information systems are available as an in-vehicle information facility, such as, CNS(Car Navigation System), HUD(Head-Up Display), RDS-TMC(Radio Data System-Traffic Message Channel), FM-DARC (Frequency Modulation-DAta Radio Channel), cellular phone, and so on. They provide the static or dynamic real-time traffic and road information to the driver via the various HMI(Human Machine Interface) modalities. However, if their HMI modalities are not properly designed and presented to the driver, they can arouse the unsafe and inefficient PIEV(Perception-Identification-Emotion-Volition) process that affects the driver and the road safety. Because they are additional information sources in addition to the basic those of instrument panel and roadside information systems(Wickens, 1980). Hence, the driver's visual activities become more disordered and sophisticated ways by distracting eye glances and durations from the road ahead. Most reported human factors evaluations of ITS-oriented information systems have been performed, which are focused on the analysis of the driver's visual activities or eye movements under the assumption that the less eye directions, frequencies and duration demanded system or a display is the more beneficial one for the driver's safety. In company with the analysis of eye movement, physiological parameters and secondary task measurement are frequently used techniques for HMI research. However, application of physiological parameters is considered quite a disappointing and heavy methodology to evaluate the driver's mental workload in a real road situation. The sophisticated instrumentations impose another experiment workloads that affect the road safety and actual driver's workload(Michon, 1993). The secondary task measurement is typically considered as a laboratory approach, taking into account the consequences in terms of interference in a real situation and the information that was obtained may be restricted(Shingledecker, 1982). Furthermore, the driving with an in-vehicle information system is already a dual task by adding a supplementary task raises questions about the driver choice in terms of priority, which task is considered the main one(Pauze et al, 1995).

These kinds of parameters are the important criteria to define the potential consequences for the disturbance of the driving task to test or to compare the system design and to establish the performance standards that can apply to any system. However, these are not sufficient criteria to assess the system usability, subjective preferences and the actual driver's workload. Moreover, human beings are usually very adaptable to the new situations, it is very difficult to analyze the human factors characteristics by these kinds of simple stimulus-response types of approaches. Therefore, most of human factors evaluations of ITS-oriented in-vehicle information systems have been performed that are focused on such objective evaluations with various subjective mental workload techniques that are regarded as the subjective evaluation or subjective rating technique, complementary, rather than conflicted one from another. They are also time and cost consuming, and because of the complexity of the driving situations and the inter-individual variability of the driver, it is certainly valuable to have an evaluation process as complete as possible(Pauze et al, 1995; Cha and Park, 1999).

Subjective mental workload assessment techniques are in correspondence with the subjective rating techniques by formalizing the driver's own judgment about the workload he or she experienced through the use of subject's judgments. They can suggest the comparative safety, subjective preferences and usability results among target systems or information displays, and then, provide the appropriate tools to draw out the HMI design guidelines and evaluation standards.

Many researches have dealt with the comparative analysis among acceptable techniques to find what is the most appropriate technique for a specific system. However, despite the increasing interest and use of them, the increasing number of technique creates both an opportunity and a problem for the human factors practitioners and researchers. On one hand, tools have been developed for a wide variety of situations; on the other hand, human factors specialists may need the valid information to make an appropriate technique choice (Hill and et al, 1992).

To suggest validated subjective mental workload assessment technique selection guideline for ITS-oriented in-vehicle information systems and other applications, this paper reviewed and compared four techniques of NASA-TLX(Hart and Staveland, 1988), SWAT(Reid and Nygren, 1998), MCH scale (Wierwille and Casali, 1983) which are receiving the greatest attention and sensitivity and having the widest range of applicability(Wierwille, 1993), and recently developed RNASA-TLX(Cha and Park, 1999). An experiment was performed between CNS-supported and non CNS-supported driving conditions in an urban area. Because the CNS is a representative ITS-oriented in-vehicle information system, which represents the current position and route guidance information through the digital map and the voice route guidance function. Comparative results of four techniques between two conditions were evaluated in terms of sensitivity and subjective evaluation criteria that was directly obtained from participants.

II. Subjective Mental Workload Assessment Techniques

1. Objectives of Subjective Mental Workload Assessment Techniques

Assessment of mental workload can help determine which tasks can be performed simultaneously with little or no decrement. Because mental workload varies as a function of the perceptual, cognitive, and motoric requirements imposed on an operator, the structure of the task and the environment in which it is performed can significantly affect on the workload and performance(Proctor and Zandt, 1994). Therefore, altering either the amount of resources available within the person or the demands made by the task on the person or group for the assessment can change the demanded mental workload of the person. Recently, application of this technique is considered as the essential step in validating and certifying the design decisions. And, they have become a standard tool for the evaluation of HMI, especially in aviation(Vidulich, 1991). They have been used on aircraft certification, aviation safety, cockpit design, aircrew evaluation, and aircraft tactical effectiveness using an operational version of the system or a simulator(many application reports are available in NASA technical report server at http://techreports.larc.nasa.gov). Recently, they have been widely applied to the vehicle and the ITS-oriented in-vehicle information systems, so that they become a standard human factors and HMI evaluation tool with following purposes(Cha and Park, 1999).

(1) Allocate the functions and tasks between the in-vehicle information system and the driver in response to the increase and decrease in predicted mental workload.
(2) Compare the imposed workloads among alternative interface components, formats or designs (for example, the color, typography or array of the display, and the information structure, HMI modality and presentation styles, and so on).
(3) Monitor the in-vehicle information system-supported driver to adapt the task difficulty or training.
(4) Evaluate the system usability, safety, and driver's subjective preferences.
(5) Maintain the driver's workload at a certain level that will allow acceptable performance.

The result of subjective technique approaches allows a typically relative comparison between situations and systems. They have been extremely popular in operational settings because of the high face validity and the ease of data collection. Furthermore, subjective measures are relatively inexpensive to obtain, nonintrusive, convenient, easy to analyze, and adaptable to different situations, so that they tend to be accepted by operators without specific machines or equipments. Although useful, there are some limitations; (1) they may not be sensitive to the aspects of the task environments that affect primary-task performance, and hence, it may be best to couple their use with primary-task measures, (2) they lack diagnosticity, (3) operators may confuse perceived difficulty with perceived expenditure of effort, (4) many factors that determine workload are inaccessible to conscious evaluation(Proctor, 1994). However, they turned out to be extremely useful in many contexts and data may still sufficiently rich in information to be useful. Reliabilities of the mood scales used in the present study have been found to range between 0.65 and 0.95(Svensson et. al, 1997).

2. Descriptions of Four Major Subjective Mental Workload Assessment Techniques for this Study

Among four considered techniques for this study, NASA-TLX is the most widely applied technique for CNS(Burnett, and Joyner, 1993; Parkes and Burnett, 1993; Pauze and et al, 1995), RDS-TMC (Vaughan, 1994), adaptive cruise control(Peter, 1996), cellular phone(Tokunaga et al, 1998), and so on. MCH scale(Parkes, 1993), RNASA-TLX(Cha and Park, 1997; 1998; 1999), and SWAT were included for this comparative study, because they are also, widely used and accepted techniques for various domains as well as ITS-oriented systems. Detailed descriptions and characteristics of four techniques are below.

1) NASA-TLX

This technique was firstly developed by the Human Performance Group at NASA Ames Research Center, and has been used on broad samples of people in various situations, and also, more recently, in the driving and ITS environments. NASA-TLX has the multidimensional rating scale procedure that uses six dimensions to assess the mental workload ; mental demand, physical demand, temporal demand, performance, effort, and frustration level as described in 〈Table 1〉. Dimensions had been developed those were based on the extensive research inclusions and psychometric analyses of laboratory experiments and simulated flight experiments. This technique is regarded as the best predictor of subjective mental workload, so that many applications have been performed on various domains, for example, on-line self assessment of air traffic control(David, 1995), evaluation of assistive technology(Stevens and Edwards, 1996), virtual interfaces(Hancock, 1996), free flight situation awareness(Endsley, 1997), auditory interface evaluation of exploration task(Pitt and Edwards, 1997), workload assessment of emergency medical staff (Jay and Morey, 1998), teleoperation system of nuclear applications (Alan, 1996), and so on.

In the first step, participants make pairwise comparisons of six dimensions of workload. This step generates weighting for each dimension, where the highest possible weighting is 5 and the lowest possible weighting is 0. The total number of weights adds up to 15, which is the number of pairwise comparisons. In the second step, subjects rate their perceived workloads on each of these individual dimensions on a scale from 0 to 100. These scores represent the raw ratings. The overall workload score for the condition is obtained by multiplying the ratings by dimension weights and by dividing the sum by 15(Hart and Staveland, 1988).

2) SWAT

This technique was developed by the U.S. Air

〈Table 1〉 NASA-TLX Six Scales and Descriptions

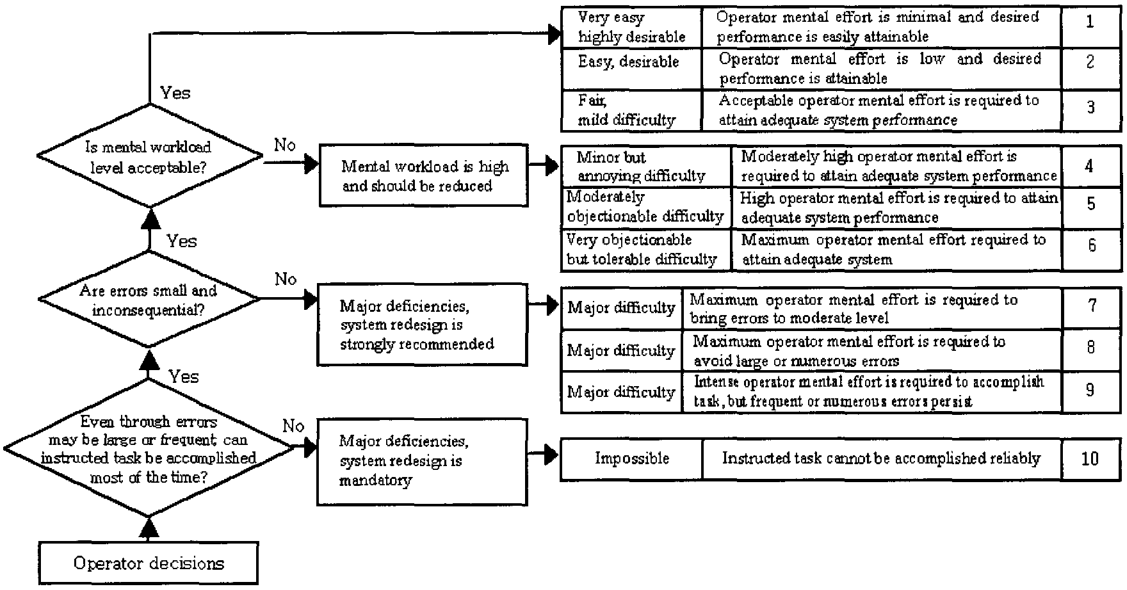
Scale	End Point	Description
Mental demand	Low/High	How much mental and perceptual activity was required(e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical demand	Low/High	How much physical activity was required(e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Performance	Low/High	How successful do you think you were in accomplishing the goals of the task set by the experimenter(or yourself)? How satisfied were you with your performance in accomplishing these goals?
Temporal demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Effort	Low/High	How hard you have to work(mentally and physically) to accomplish your level of performance?
Frustration level	Low/High	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

〈Table 2〉 Three-Point Rating Scales of SWAT

Time Load	Mental Effort Load	Stress Load
1. Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.	1. Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.	1. Little confusion, risk, frustration, or anxiety exists accommodated.
2. Occasionally have space time. Interruptions or overlap among activities occur frequently.	2. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention required	2. Moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.
3. Almost never have space time. Interruptions or overlap among activities are very frequent or occur all the time.	3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.	3. High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required.

Force Armstrong Aerospace Medical Research Laboratory at Wright-Patterson Air Force Base to provide a psychological model of subjective mental workload that is based on an additive multidimensional representation of three dimensions of time load, mental effort load and stress load for the use of various tasks and systems. Time load refers to the extent to which a task must be performed within a limited amount of time and the extent to which multiple tasks must be performed at the same time. Mental effort load involves inherent attentional demands of tasks, such as attending to multiple sources of information and performing calculation. Stress load encompasses operator variables, such as fatigue level of training, and emotional state, which contribute to an operator's anxiety level and each description is in 〈Table 2〉 (Procter and Zandt, 1994).

Evaluation procedure is divided into two distinct steps. In the first step, the scale development phase, subjects are asked to order all 27 possible combinations of the three descriptions according to their amount of workload in a rank order to assess the validity of fitting workload judgment



〈Figure 1〉 The Modified Cooper-Harper

to an additive model. Typically, operators seem to add the difficulty of the separate task descriptions to arrive at estimates for the difficulty of the combinations. Then, the conjoint measurement is used to develop the interval scale of actual workload rating from 0 to 100 for a given task. In the second step, the event scoring phase, an activity or event is rated by assigning a value of 1 to 3 on each of the three dimensions(Nygren, 1985; 1986). The scale value associated with this combination(obtained from the scale development phase) is then assigned as the workload value for that activity. Given such a correspondence, the event scores can be translated into individual component scores on the time, mental effort, and stress load dimensions and simultaneously to an additional overall workload score. The SWAT procedure is sensitive to workload increases induced by increase in task difficulty, as well as to those caused by sleep deprivation or increased time-on-task(Hankey and Dingues, 1990).

3) MCH Scale

The original CH(Cooper-Harper) scale is directed at the rating of aircraft handling qualities and as such does not lend itself well to ratings of more general workload dimensions. Wierwille and Casali (1983) have modified the CH verbal descriptors, MCH scale. This is still ordinal, to be applicable to a wider variety of workload evaluation environments including tasks with a communication emphasis(Casali, 1983). MCH has a 10-point unidimensional rating scale that results in a global rating scale of workload. The rating scale uses a decision tree to assist the rater in determining the most appropriate rating to assign between 1(low workload) and 10(high workload) that is illustrated in 〈Figure 1〉. These numbers are best conceived as ordinal indicators of the degree of mental workload(Proctor and Zandt, 1994). MCH scale has been shown to be sensitive to differences in workload and to be consistent across tasks, despite the fact that it was not derived psychometrically(Skipper and et al, 1986).

4) RNASA-TLX

Although NASA-TLX is widely acceptable as a valid subjective mental workload assessment, its predictive validity is uncertain and its use as psychological models of workload is questionable (Nygren, 1991). In addition, because NASA-TLX was originally designed for the evaluation of a pilot and the aviation tasks that require very specific training and a level of education very often higher than the average person, previous experimental results have shown a wide range of variability between subjects(Marin and Dejeamme, 1995). And, this measurement has shown the unsuitability for the impact evaluation of an in-vehicle information system to the driver, because of translation problem and inefficient descriptions of operations and control status(Pauze et al, 1995; Cha and Park, 1997).

RNASA-TLX was firstly designed to develop the robust and understandable scales for the evaluation of in-vehicle information systems on the basis of NASA-TLX. During RNASA-TLX development process by conducting the experiment to evaluate the impact of CNS use, most of subjects faced with the difficulties in understanding NASA-TLX six dimensions despite the detailed explanations and the training sessions before experiments(Cha and Park, 1997). It is because NASA-TLX six dimensions consist of technical, vague, and unfamiliar words that do not contain any specific words, which were related to operation, control and HMI between CNS and driver. Also, subjects required the well-defined and clear meanings of descriptions to express and rating his or her workload successfully.

Firstly, HMI objects analysis of the real products was performed and its results were reflected to the RNASA-TLX by selecting and prioritizing HMI design variables from the experts and users(Cha and Park, 1997). In addition, Pauze et al's(1995) NASA-TLX evaluation results were considered that pointed out following problems; (1) physical demand does not consider the vehicle automation trend, (2) mental workload contains the perceptive and cognitive aspects of the workload that is deeply related with the context of the driving task to identify each of various modalities, (3) performance does not consider any objective data. Then, subjects' opinions and complaints had been reflected during experiments and evaluation stages in a real road and the laboratory simulator experiments, for example, information modality comparison(Cha and Park, 1997), digital map presentation method comparison(Cha and Park, 1998). As results, six dimensions of NASA-TLX had been changed more distinctly to minimize the correlation among six dimensions for the better understandings of the subjects and for clear results.

RNASA-TLX described in 〈Table 3〉, and its six dimensions include all considerable mental workload sources of an in-vehicle information system and take the useful aspects of NASA-TLX. For example, mental demand includes the overall mental workload sources of driving with the in-vehicle information system, whereas, that of NASA-TLX contains performance-related factors of driving without any objective data. Temporal demand contains all kinds of factors that have the possibility to occur the time pressure. Also, description of difficulty in driving refers the detailed vehicle-related equipments. In addition, other dimensions contain the visual, auditory, and information-related factors to refine all kinds of mental workload sources of the in-vehicle information system.

Execution procedure has two-step procedures like NASA-TLX. In a weighting procedure, the tallies had been changed into 5, and Saaty's(1980) 'rule of thumb' was used. So, when judgments are inconsistent, the experimenter could give the opportunities to revise the pairwise comparisons. Then, subjects rate six dimensions using scales of 0 to 100. Several applications have been performed, for example, manual vehicle driver's operation performance(Yu, 1998), comparison of vehicle IP design(Park, 1998), workload comparison in virtual driving environment(Park, 1999), and fidelity test of CNS simulator(Cha and Park, 1999), and so on.

〈Table 3〉 RNASA-TLX Six Scales and Descriptions

Scale	End Point	Description
Mental demand	Low/High	How much mental attention was needed during driving when using in-vehicle information system? Namely, how much mental stress was required during driving via in-vehicle information system : to keep the lane, to avoid the collision, to observe the traffic law, to find the route, and so other things, which related to driving activity.
Visual demand	Low/High	How much visual activity was required during driving when using in-vehicle information system to recognize the information from in-vehicle information system or other external information sources? For example, digital map and its information and the traffic signal, rear mirror, variable message signs and so other external information sources.
Auditory demand	Low/High	How much auditory activity was required during driving when using in-vehicle information system to recognize or hear the information presented from in-vehicle information system or other auditory sources?
Temporal demand	Low/High	How much time pressure was required due to rate or pace at the task elements occurred during driving using in-vehicle information system? For example, in operating or menu selection process and information presentation pace or speed.
Difficulty in driving	Low/High	How hard you driving when using in-vehicle information system with other in-vehicle control equipment or optional devices? For example, cellular-phone, gear stick, side break, audio, and so on?
Difficulty in understanding information	Low/High	How hard you understanding the information presented from in-vehicle information system? Was the information from in-vehicle information system compatible with your association? Was the mass, density? And other information related factors are suitable for you?

Ⅲ. Experiment Design

An experiment was divided into two conditions for more accurate and comparative workload expressions between a CNS-supported driving and non CNS-supported conservative driving. Experiment vehicle was the manual type with the CNS that was mounted at the middle of the dashboard on the rightside of the driver. An equipped CNS has the 5-inch LCD color display that presents the digital map and voice route guidance by remote controller operation with the north-up map presentation. Descriptions of each condition are below.

(1) Experiment condition 1 : CNS-supported driving (V-100 experiment vehicle of D motors company with the CNS of H electronic company)
(2) Experiment condition 2 : Non CNS-supported driving(V-100 experiment vehicle of D motor company)

Experiments were performed in daylight at Suwon. 10 demographically homogeneous group of male subjects were participated whose age was from 26 to 28 years old(average : 27.6, s.d. : 0.52), and 5.8 years of average actual driving experiences. They were the novices of the CNS, and the commuters from other cities to Suwon. Therefore, they were not accustomed to the Suwon area in detail, but they were to their own commuting ways. Before experiments, each subject was requested to select the unknown route and destination to which he had to drive as the experimental routes about 30 to 60 minutes distance. All subjects were volunteers and were paid a nominal gratuity for participation. Also, detailed workload concept and evaluation procedures were educated to each subject.

In condition 1, subject had to reach the predetermined destination mostly depended on the CNS information, and in condition 2, they had to drive the vehicle mostly depended on the roadside direction signs, pedestrians or other driver's consulting, and other conservative wayfinding methods. Evaluation of subjective mental workload had been performed just after each experimental driving using evaluation sheets without time limitations to express their imposed workload efficiently and accurately.

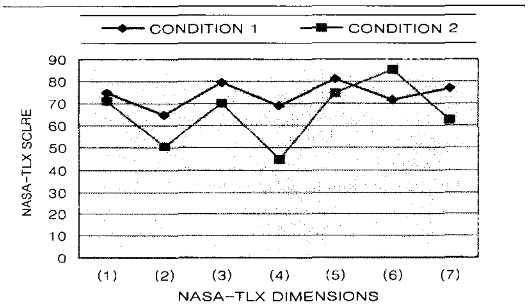
Ⅳ. Experiment Results

1. Experiment Results of Four Subjective Mental Workload Techniques

⟨Figure 2⟩ shows the NASA-TLX results between two conditions. Result indicates that condition 1 required the higher mean weighted workload score than condition 2. It means that the drivers felt and received the increased workload when CNS-supported driving in an unknown urban road networks. However, all six dimensions were not showed statistically significant differences by Wilcoxon non-parametric one-tailed test. Also, as illustrated in ⟨Figure 3⟩, RNASA-TLX result shows the higher all six dimensions and mean weighted workload score in condition 1. Except difficulty in driving ($W_{observation}=14$, $W_{critical\ value}=8$, when $n=10$, $\alpha=0.05$), other dimensions showed the significant differences by Wilcoxon nonparametric one-tailed test. By these results, it is certain that a CNS requires more driver mental workload when driving unknown destination in an urban area. However, RNASA-TLX suggested more statistically validated result than NASA-TLX. Also, RNASA-TLX explains the more detailed and entire mental workload sources that are related to the driver's information acquisition process and HMI objects.

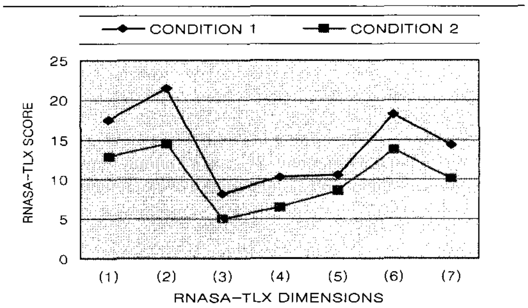
In ⟨Figure 4⟩, MCH result of the experiment condition 1 is referring 'major difficulty' and condition 2 is 'very objectionable but tolerable difficulty'. This result showed the significant differences by Wilcoxon non-parametric one-tailed test($W_{observation}=5$, $W_{critical\ value}=5$, when $n=9$,(=0.05). However, this result just explained the comparative evaluation between two conditions without any detailed system usability and subjective information.

⟨Figure 5⟩ represents the SWAT result, which shows condition 1 requires higher time and mental



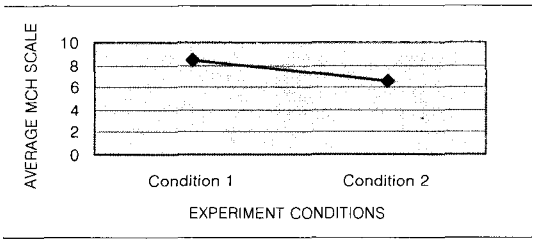
(1) Mental demand (2) Physical demand (3)Temporal demand (4) Performance (5) Effort (6) Frustration Level (7) Global weighted workload score

⟨Figure 2⟩ Results of NASA-TLX

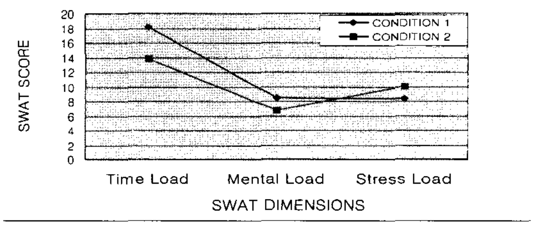


(1) Mental demand (2) Visual demand (3) Auditory demand (4) Temporal demand (5) Difficulty in driving (6) Difficulty in understanding information (7) Mean weighted workload score

⟨Figure 3⟩ Results of RNASA-TLX



⟨Figure 4⟩ Results of MCH Scale



⟨Figure 5⟩ Results of SWAT

〈Table 4〉 Example Procedure of SWAT of Experiment Condition1(white area) and Results of Each Condition (gray area)

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10
Scale value										
Time load 1	0.8594	-0.7758	-0.7081	-1.0129	0.6313	-0.8260	-0.8177	0.0789	-0.2721	-0.8897
Time load 2	0.1693	-0.1395	-0.1395	-0.1996	0.1244	-0.1538	-0.1621	0.0155	-0.0536	-0.1753
Time load 3	-1.0287	0.9153	0.8476	1.2125	-0.7557	0.9798	0.9798	-0.0944	0.3257	1.0650
Mental load 1	-0.5331	-0.5470	-0.0371	-0.4425	0.0644	-0.5334	-0.4425	0.0489	-0.4338	0.4668
Mental load 2	-0.0787	-0.0808	-0.0055	-0.0655	0.0095	-0.0787	-0.0653	0.0072	-0.0640	0.0689
Mental load 3	0.6118	0.6278	0.0426	0.5080	-0.0739	0.6124	0.5080	-0.0561	0.4981	-0.5357
Stress load 1	-0.0231	-0.0824	-0.0792	-0.0479	-0.3439	-0.0127	-0.0479	-0.3439	-0.0980	-0.7081
Stress load 2	-0.0609	-0.2176	-0.2091	-0.1263	-0.9077	-0.0335	-0.1263	-0.9077	-0.2588	-0.1395
Stress load 3	0.0840	0.3000	0.2883	0.1742	1.2516	0.0462	0.1742	1.2516	0.3568	0.8476
Weights (scale development phase)										
Time Load	0.5941	0.4998	0.7294	0.6401	0.3764	0.5957	0.5896	0.0711	0.2787	0.4331
Mental Load	0.3603	0.3472	0.0374	0.2734	0.0375	0.3780	0.3118	0.0431	0.4344	0.2221
Stress Load	0.0456	0.1530	0.2332	0.0864	0.5860	0.0263	0.0986	0.8858	0.2870	0.3447
Three dimension rating (event scoring phase)										
	M〉T〉S	M〉T〉S	M〉S〉T	T〉M〉S	T〉M〉S	T〉M〉S	T〉S〉M	T〉S〉M	S〉T〉M	S〉M〉T
SWAT results of two conditions										
SWAT score of condition1 (CNS-supported driving)										
Time load	17.6458	14.8443	11.9624	34.5014	20.2880	32.1783	31.7741	3.8323	8.2761	7.1034
Mental load	19.4185	18.7149	2.0142	8.1200	1.1167	11.1910	9.2634	0.7068	7.1240	6.5976
Stress load	0.7478	2.5088	6.9265	1.4170	9.6104	0.4297	1.6170	26.3083	15.4668	18.5806
SWAT score of condition 2 (non CNS-supported driving)										
Time load	10.3742	10.2821	4.8300	17.4120	14.8112	10.3125	22.3130	21.4944	12.3791	13.8278
Mental load	5.2789	12.5643	9.3604	10.811	2.9183	5.2972	5.1340	3.6226	4.9776	8.0710
Stress load	5.6271	11.6038	7.939	3.365	15.5590	4.9147	3.9147	16.3921	8.2171	23.2710

load, but condition 2 requires more stress load. 〈Table 4〉 shows the SWAT procedures of condition 1(white area) and stepwise result of each condition (gray area).

Then, Wilcoxon nonparametric one-tailed test results indicated that three dimensions did not show the significant difference between two experiment conditions(time load($W_{observation}=15$, $W_{critical\ value}=5$, when $n=9$, $\alpha=0.05$), mental load ($W_{observation}=21$, $W_{critical\ value}=5$, when $n=9$, $\alpha=0.05$), stress load ($W_{observation}=18$, $W_{critical\ value}=5$, when $n=9$, $\alpha=0.05$). This result suggested more specific evaluation information than MCH, however, it is no more than a comparative result between two experiment conditions. SAS release 6.12 and Microsoft™ Excel had treated all statistical procedures and results obtained for this experiment.

2. Comparative Analysis of Four Subjective Mental Workload Assessment Techniques

To examine how each of four techniques was able to discriminate between two experiment conditions, sensitivity of each technique was measured using two-stage factor validity analysis that was proposed by Hill et al(1992). In the first stage, a principal component analysis is conducted on the sets of segment ratings collected across subjects and two experiment conditions. Each set included workload ratings of four techniques. Then, the

〈Table 5〉 Factor Validity of Four Techniques (α=0.05)

Experiment condition	% Total variance	Factor Validates			
		NASA-TLX	RNASA-TLX	MCH scale	SWAT
Condition 1	80%	0.907	0.915	0.862	0.824
Condition 2	77%	0.896	0.923	0.858	0.837

〈Table 6〉 Subjective Evaluations of Four Techniques by Participants

	NASA-TLX	RNASA-TLX	MCH scale	SWAT
Subjects acceptance	2.2	2.1	1.6	3.9
Fitness of evaluation	2.1	1.5	2.4	3.5

* Average of mean rankings (1=best .... 4=worst) of 10 subjects

results of this initial analysis supported the view that four workload techniques essentially provide assessments of a single common factor of driver's workload factor. Jackknife principal component analyses were conducted of the workload measures during the second stage in order to evaluate the stability of the factor loading of the four techniques. Jackknife analysis generally involves successive analysis, dropping subjects one at a time from a data set in order to allow analysis of the stability of parameter estimates(Hinkley, 1983). And then, the Analysis of Variance was used to examine significant differences among the workload technique factor loading. Results are shown in 〈Table 5〉, RNASA-TLX obtained the highest factor validity for each of two experiment conditions, then, NASA-TLX, MCH, and SWAT followed in two conditions. These results showed all statistically significant differences between factor validities(α=0.05). Another source of comparative results of the four mental workload techniques is the subjective evaluation directly obtained from the participants. These dimensions are very important because the increased driver's acceptance of the workload technique may result in increased willingness to express a valid opinion that can be taken seriously and used. First dimension is subject's acceptance, which means the easiness of evaluation procedure and completion. The other is the fitness of evaluation that means 'what technique is composed with the most understandable and suitable contents for the evaluation workload imposed from in-vehicle information system?'. 〈Table 6〉 shows the results of these subjective evaluations. Results show that MCH scale obtained the highest acceptance and RNASA-TLX obtained the highest acceptance in fitness of evaluation.

V. Conclusion and Future Enhancement

By the development of the mobile communication, computer and vehicular technologies, several kinds of in-vehicle systems are available for commercial drivers as well as common drivers. Human factors approach regarded as an essential procedure when designing and evaluating those kinds of systems for the driver-centered HMI design to improve the system usability and safety. Among various human factors evaluation techniques, subjective mental workload assessment technique is a useful and economic one by reducing the time and cost for experiment and by supplying the directly obtained driver's preferences and opinions without evaluation equipments.

This paper reviewed and compared four widely accepted subjective workload assessment techniques for the application of in-vehicle information systems and other ITS-oriented applications. Despite of small subjects size, the results indicated important implications for the use of subjective mental workload assessment technique for ITS-oriented human-machine systems. The results support that RNASA-

TLX is the most sensitive and acceptable technique when performing the mental workload assessment of CNS, and MCH obtained the highest acceptance, because of its diagram type of evaluation scheme. But, MCH scale only can suggest the simple comparative evaluation result without specific information of workload sources and usability. Also, the results imply the importance of careful selection and revision process of subjective mental workload assessment technique by considering the HMI characteristics of target evaluation system and operation environments.

RNASA-TLX could be applicable as a subjective mental workload assessment technique for other various ITS-oriented in-vehicle and roadside information systems, for example, RDS-TMC, FM-DARC, HUD, flat panel display of Advanced Highway System, variable message signs, cellular phone, and others, which have the similar HMI characteristics with CNS. In the evaluation process of those kinds of systems, if cautiously selected objective evaluation techniques which considering the system HMI characteristics combined with RNASA-TLX, the accuracy and implications of the results would be improved in terms of Korean drivers. However, because the results of this study were extracted only from a CNS-based driving evaluation procedure, more experiments by considering various system and operation environments were required for the development of more validated subjective mental workload assessment techniques.

References

1. Alan, A. B., An interface design for a teleoperation system for nuclear applications, *Working Paper*, Texas Universities Robotics and Automation Consortium, http://www.osci.ttu.edu/TURAC/Interface-dgn/index.html, 1996.
2. Casali, J. G. and Wierwille, W. W., Comparison of rating scale, secondary-task, physiological and primary-task workload estimation techniques in a simulated flight task emphasizing communication load, *Human Factors*, 26(6), pp.623~641, 1983.
3. Casali, J. and Wierwille, W. A., On the measurement of pilot perceptual workload : A comparison of assessment techniques addressing sensitivity and intrusion issues, *Ergonomics*, 27 (10), pp.1033~1050, 1984.
4. Cha, D. W. and Park, P., User required information modality and structure of in-vehicle navigation system focused on the urban commuter, *Computers ind Engng*, 33(3-4), pp.517~520, 1997.
5. Cha, D. W. and Park, P., Simulator-based mental workload assessment of the in-vehicle navigation system driver using Revision of NASA-TLX, *IE-Interfaces*, 10(11), pp.145~154, 1997.
6. Cha, D. W. and Park, P., Development of human-machine interface assessment system and derivation of design variables for vehicle navigation system, *Journal of Korean Society of Transportation*, 15 (2), pp.35~48, 1997.
7. Cha, D. W. and Park, P., Ergonomic research of the map-rotation and the text-rotation method of the navigation system digital map, *Journal of Korean Society of Transportation*, 16(1), pp.47~58, 1998.
8. Cha, D. W. and Park, P., Human sensitivity ergonomics evaluation for car navigation system digital map, *Journal of the Society of Korea Industrial and Systems Engineering*, 21(48), pp.101~111, 1998.
9. Cha, D. W. and Park, P., Cognitive evaluation of car navigation system digital map prototypes using drawing method, *Journal of Traffic Safety Research*, 17, pp.125~137, 1998.
10. Cha, D. W. and Park, P., Development human factors evaluation system for car navigation system, *IE-Interfaces*, 12 (2), pp.294~304, 1999.
11. David, H. S. P., Intrusiveness of on-line self-assessment in ATC simulation using keyboard and speech recognition, *EEC Reports NO. 275*, European Organization for the Safety of Air Navigation, 1995.
12. Endsley, M. R., Situation awareness, automa-

tion & free flight, *Proceeding of 1st U.S.A./ EUROPE Air Traffic Management Seminar*, Saclay, pp.17~20 June, 1997.
13. Guy, L., Influence of in car navigation map displays on drivers performance, *SAE Paper No. 891683*, 1989.
14. Hancock, P. A., Effects of control order, augmented feedback, input device and practice on tracking performance and perceived workload, *Ergonomics*, 39(9), pp.146~162, 1996.
15. Hankey, J. M. and Dingues, T. A., A validation of SWAT as a measure of workload induced by changes in operator capacity, *Proceeding of the Human Factors Society 34th Annual Meeting*, 112-115, Santa Monica, 1990.
16. Hart, S. G., and Staveland, L., Development of NASA-TLX(Task Load Index) : Results of empirical and theoretical research, In P. A. Hancock and N. Meshtati(Eds.), *Human mental workload*, Amsterdam, Elsevier, 1988.
17. Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C., Zaklad, A. L. and Christ, R. E., Comparison of four subjective workload rating scales, *Human Factors*, 34(4), pp.429~439, 1992.
18. Hinkely, D. V., Jackknife method, In S. Konz, N. L. Johnson, and C. B. Read(Eds.), Encyclopedia of statistical science, 4, pp.280~287, New York, Wiley, 1983.
19. Jay, G. D. and Morey, J., ED staff workload-subjective and objective measures, *Proceeding of Society for Academic Emergency Medicine 1998 Annual Meeting*, pp.17~20, Chicago, 1998.
20. Lynn A. S., A profile of drivers' map-reading abilities, *Human Factors*, 28(2), pp.223~240, 1986.
21. Marin, L. C. and Dejeamme, M., The processing of complex guidance symbols by elderly drivers : A simulator-based study and an evaluation of CARMINAT guidance system by European Community DRIVE-EDDIT project, *Proceeding of IEEE-IEE Vehicle & Information Systems Conference*, Seattle, pp.114~117, 30 July-2 August, 1995.
22. Michon, J. A., *Generic intelligent driver support*, Taylor & Francis, 1993.
23. Nygren, T. E., An examination of conditional violations of axioms for addictive conjoint measurement, *Applied Psychological Measurement*, 9, pp.249~264, 1985.
24. Nygren, T. E., A two-stage algorithm for assessing violations of additivity via axiomatic and numerical conjoint measurement, *Psychometrika*, 51, pp.483~491, 1986.
25. Nygren, T. E., Psychometric properties of subjective workload measurement techniques : Implications for their use in the assessment of perceived mental workload, *Human Factors*, 31(1), pp.17~33, 1991.
26. Park, P., Study on the guideline and development of Human Sensibility Engineering model for car interior design, *Technical Report*, Industrial Engineering Department of Ajou University, 1998.
27. Park, P., A development of the model and the prototyping interface for human-machine system interactions, *Technical Report, Industrial Engineering Department of Ajou University*, 1999.
28. Parkes, A. M. and Burnett, G. E., An evaluation of medium range advance information in route-guidance displays for use in vehicle, *Proceeding of IEEE-IEE Vehicle & Information Systems Conference*, pp.238~241, Seattle, 30 July-2 August, 1993.
29. Pauze, A., Shrpedon, A. and Samulnier, G., Ergonomic evaluation of a prototype guideline system in an urban area : Discussion about methodologies and data collection tools, *Proceeding of IEEE-IEE Vehicle Navigation & Information Systems Conference*, pp.390~396, Seattle, 30 July-2 August, 1995.
30. Peter, B., Evaluation of an adaptive cruise control(ACC) system used by drivers with lower limb impairments, *Proceeding of 3rd Annual World Congress on Intelligent Transport Systems*

(CD-ROM Version), Orlando, 14-18 Oct, 1996.

31. Pitt, L. J. and Edwards, A. D. N., An improved auditory interface for the exploration of lists, *ACM Multimedia 97-Electronic Proceedings*(http : //www.uni-mannheim.de/acm97/papers/pitt/paper.html), Seattle, 8-14 November, 1997.
32. Proctor, R. W. and Zandt, T. V., *Human factors in simple and complex systems*, Allyn and Bacon, 1994.
33. Reid, G. B. and Nygren, T. E., The subjective workload assessment technique : A scaling procedure for measuring mental workload. In P. A. Hancock and N. Meshtati(Eds.), *Human Mental Workload*, Amsterdam, Elsevier, 1998.
34. Rockwell, T. H., Spare visual capacity in driving - revisited, *Proceeding from Vision in Vehicle III Conference*, Aachen, Germany, 1989.
35. Saaty, T. L., *The analytic hierarchy process, New York*, MCgRAW-Hill, 1980.
36. Svensson, E., Thanderz, M. A., Sjoberg, L. and Olsson, S., Information complexity-mental workload and performance, *Combat Aircraft, Ergonomics*, 40(3), pp.362~380, 1997.
37. Shingledecker, C. A., Enhancing operator acceptance and noninterference in secondary task measures of workload, *Proceeding of the 24th Annual Meeting of the Human Factors Society*, pp.674~677, Santa Monica, 1980.
38. Shingledecker, C., Performance evaluation of the embedded secondary task technique, *Aerospace Medical Association Annual Scientific Meeting*, pp.151~152, 1982.
39. Skipper, J. H., Rieger, C. A. and Wierwiller, W. W., Evaluation of decision-tree rating scales for mental workload estimation, *Ergonomics*, 29, pp.383~399, 1986.
40. Stevens, R. D. and Edwards, A. D. N., An approach to the evaluation of assistive technology, *Proceedings of Assets 96*, pp.64~71, Vancouver, 1996.
41. Tokunaga, R., Hagiwara, T., Takagi, H. and Shimojo, A., Effects of cellular telephone use on driver′s reaction time and subjective mental workload, *Proceeding of 5th World Congress on Intelligent Transport Systems(CD-ROM Version)*, Seoul, 1998.
42. Vaughan, G., A Human factors investigation of an RDS-TMC system, *Proceedings of 1st World Congress of ITS*, pp.1685~1692, Paris, 30 November-3 December, 1994.
43. Vidulich, M. A., Using the subjective workload dominance technique for projective workload assessment, *Human Factors*, 33(6), pp.677~691, 1991.
44. Wickens, C. D., The structure of attentional resources, In R. Nickerson & R. Pews(Ed.), *Attention and Performance*, Englewood Cliffs, NJ, Prentice Hall, 1980.
45. Wierwille, W. W. and Casali, J. G., A validated rating scaling for global mental workload measurement applications, *Proceedings of the Human Factors Society 27th Annual Meeting*, pp. 129~133, Santa Monica, 1983.
46. Wierwille, W. W. and Eggemeier, F. T., Recommendations for mental workload measurement in a test and evaluation environment, *Human Factors*, 35 (2), pp.263~281, 1993.
47. Yu, S. D., A study of the relationship between the time to implement of driving secondary task and mental workload to standardness of cockpit, *Master Dissertation, Industrial Engineering Department of Ajou University*, 1998.

♧ 주 작 성 자 : 차두원
♧ 논문투고일 : 2001. 4. 9
논문심사일 : 2001. 6. 7 (1차)
2001. 6. 12 (2차)
심사판정일 : 2001. 6. 12