Applied Ergonomics 60 (2017) 260–274

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

A colored petri nets based workload evaluation model and its validation through Multi-Attribute Task Battery-II

Peng Wang, Weining Fang*, Beiyuan Guo

State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, No. 3 Shang Yuan Cun, Haidian District, 100044 Beijing, China

ARTICLE INFO

Article history:
Received 8 February 2016
Received in revised form
21 November 2016
Accepted 27 November 2016

Keywords:
Colored petri nets
Workload
Multi-Attribute Task Battery-II

ABSTRACT

This paper proposed a colored petri nets based workload evaluation model. A formal interpretation of workload was firstly introduced based on the process that reflection of petri nets components to task. A petri net based description of Multiple Resources theory was given by comprehending it from a new angle. A new application of VACP rating scales named V/A-C-P unit, and the definition of colored transitions were proposed to build a model of task process. The calculation of workload mainly has the following four steps: determine token's initial position and values; calculate the weight of directed arcs on the basis of the rules proposed; calculate workload from different transitions, and correct the influence of repetitive behaviors. Verify experiments were carried out based on Multi-Attribute Task Battery-II software. Our results show that there is a strong correlation between the model values and NASA -Task Load Index scores (r=*0.9513*). In addition, this method can also distinguish behavior characteristics between different people.

1. Introduction

The evaluation of workload, which is a key parameter in the design and evaluation of human-machine interfaces, has an important effect on the comfort, satisfaction, efficiency, and safety of the workplace (Rubio et al., 2004). Previous studies have noted that level of workload is related to accident frequency (Lei et al., 2009) and job performance (Brookhuis et al., 2009). In the conference, 'Workload: Its Theory and Measurement,' in 1977, researchers taught that workload was a multi-dimensional concept: it involves work requirements, time pressure, ability of the operator, behavioral expression, and many other factors (Moray, 2013). After decades of research and development, the definition of workload has changed and is now more interpreted in terms of demanded resources on time and cognition when performing a task (Pickup et al., 2010; Shneiderman, 2012).

Methods for evaluating workload can be divided into three types: subjective self-report measures, performance measures, and physiological measures (Silvada, 2014). Subjective self-report measure are implemented by questionnaires or interviews, such as the NASA-Task Load Index (NASA-TLX) (Hart and Staveland, 1988), Subjective Workload Assessment Technology (Reid and Nygren, 1988), Multiple Resources Questionnaire (Boles and Adair, 2001), and Integrated Workload Scale (Pickup et al., 2005). Performance measures employ two main types of measurement, primary task measures (Bromfield and Dillman, 2015)and secondary task measures (Pauzié and Manzano, 2007), and use time or error rate as the calculation basis to evaluate workload. Physiological measures evaluate signals of the peripheral and central nervous system of participants, and their major measurements include Heart Rate Variability (DiDomenico and Nussbaum, 2011), Electroencephalography (Zarjam et al., 2013), and Eye Movement (Brookhuis and de Waard, 2010). Nevertheless, several researchers think that workload has a multidimensional indicators, a single evaluation method cannot reflect all aspects of workload. Such researchers have proposed a comprehensive evaluation model to solve the problem, including Overall Workload Level (OWL) (Jung and Jung, 2001)and ACT-R cognitive architecture (Jo et al., 2012). OWL examines four variables: physical job demand, environmental factors, postural discomfort, and mental job demand. In one particular study utilizing OWL, 28 workers gave a mark to each variable using a five-point value, and the analytic hierarchy process (AHP) was then applied to estimate the external workload. ACT-R cognitive architecture was originally used to evaluate operators' performances, and workload was measured by an algorithm that

* Corresponding author.
E-mail addresses: c2t53y48@gmail.com (P. Wang), wnfang@bjtu.edu.cn (W. Fang), byguo@bjtu.edu.cn (B. Guo).

http://dx.doi.org/10.1016/j.apergo.2016.11.013
0003-6870/

calculated the activated time of the ACT-R modules.

To focus exclusively on the workload of helicopter pilots, the VACP rating scales method was proposed (McCracken and Aldrich, 1984). VACP has four parts: visual demand, auditory demand, cognitive demand and psychomotor demand. Every part defines several abstract behaviors that are encoded (shown in Table 1), and the ratings range from 0 to 7. The main advantage of the VACP rating scales method was that it measured the demand on resources from basic abstract behaviors. Therefore, workload could be predicted after decomposition of the task to the behavioral level. Some researchers (Hamilton and Clarke, 2005; Wang and Fang, 2014) have used the VACP rating scales to evaluate codes of behavior in driver's performance, modeling and error prediction. As a standard of measure (Keller, 2002), VACP rating scales were applied to predict the workload of driving a car while talking on a cell phone, and the evaluation of the workload was performed by calculating the maximum, mean, and standard deviation of the task's total demand for resources after analysis of the task. To verify the effectiveness of VACP rating scales (Liang et al., 2014), a comparison was performed between the predicted results of VACP ratings scales and a NASA-TLX questionnaire that were given in a study that measured the workload of physical therapists. The task analysis was based on the physical therapists' standard operating procedures and discussion with the therapists. The correlation between the VACP ratings scales and NASA-TLX questionnaire was higher than 0.7. Although this method has been widely used, some issues should be noted:

- First, the methods for applying the VACP rating scales were not unified. Workload calculation can be divided into two steps: Determine the behaviors of the tasks according to VACP rating scales, and then calculate them through a set of rules. Expert discussion guides the first step, while different sets of rules have been proposed for the second step. For instance, one such set of rules suggested determining the behaviors in different phases of the task, and then adding the values in four dimensions: visual, cognitive, auditory and psychomotor (Liang et al., 2014); however, this method was completely subjective and ignored the influence of time. Another set of rules considered the conflict of concurrent interface channels (North and Riley, 1989). Finally, another set of rules proposed calculating workload by time occupancy ratios (Popkin et al., 2001). Some researchers used formal tools to analyze the tasks that they do not evaluate through the behavioral aspects (McIlroy and Stanton, 2015).
- Second, continuous repetitive and dispersive repetitive behaviors have different influences on the evaluation of workload. This finding was made obvious in certain behaviors that focused on interaction with the same object. For instance, consider an experiment we did that involved the repetitive task of pressing a button. The operator needed to press the button 3 times in the task procedure at different time intervals. The behaviors we recorded were three psychomotor P2 (shown in Table 1), and we added these resources demanded to the other behaviors to predict the workload. However, if the procedure involved pressing the button three times continuously (as opposed to at different time intervals), the resources demanded were less

Table 1
VACP rating scales.

Rate	Description
Visual	
0.0	No visual activity (V0)
1.0	Visually register/detect (detect occurrence of image) (V1)
3.7	Visually discriminate (detect visual differences) (V2)
4.0	Visually inspect/check (discrete inspection/static condition) (V3)
5.0	Visually locate/align (selective orientation) (V4)
5.4	Visually track/follow (maintain orientation) (V5)
5.9	Visually read (symbol) (V6)
7.0	Visually scan/search/monitor (continuous/serial inspection, multiple conditions) (V7)
Auditory	
0.0	No auditory activity (A0)
1.0	Detect/register sound (detect occurrence of sound) (A1)
2.0	Orient to sound (general orientation/attention) (A2)
4.2	Orient to sound (selective orientation/attention) (A3)
4.3	Verify auditory feedback (detect occurrence of anticipated sound) (A4)
4.9	Interpret semantic content (speech) (A5)
6.6	Discriminate sound characteristics (detect auditory differences) (A6)
7.0	Interpret sound patterns (pulse rates, etc.) (A7)
Cognitive	
0.0	No cognitive activity (C0)
1.0	Automatic (simple association) (C1)
1.2	Alternative selection (C2)
3.7	Sign/signal recognition (C3)
4.6	Evaluation/Judgment (consider single aspect) (C4)
5.3	Encoding/Decoding, recall (C5)
6.8	Evaluation/Judgment (consider several aspects) (C6)
7.0	Estimation, calculation, conversion (C7)
Psychomotor	
0.0	No psychomotor activity (P0)
1.0	Speech (P1)
2.2	Discrete actuation (button, toggle, trigger) (P2)
2.6	Continuous adjustment (flight control, sensor control) (P3)
4.6	Manipulation (P4)
5.8	Discrete adjustment (rotary, vertical thumb wheel, lever position) (P5)
6.5	Symbolic production (writing) (P6)
7.0	Serial discrete manipulation (keyboard entries) (P7)