

Towards cognitive-aware multimodal presentation: the modality effects in high-load HCI

Yujia Cao, Mariët Theune and Anton Nijholt

Human Media Interaction Group, University of Twente,
P.O. Box 217, 7500AE, Enschede, The Netherlands
{y.cao, m.theune, a.nijholt}@utwente.nl

Abstract. In this study, we argue that multimodal presentations should be created in a cognitive-aware manner, especially in a high-load HCI situation where the user task challenges the full capacity of the human cognition. An experiment was conducted to investigate the cognitive effects of modalities, using a high-load task. The performance measurements and subjective reports consistently confirm a significant modality impact on cognitive workload, stress and performance. A relation between modality usage and physiological states was not found, due to the insufficient sensitivity and individual differences of the physiological measurements. The findings of this experiment can be well explained by several modality-related cognitive theories. We further integrate these theories into a suitability prediction model, which can systematically predict how suitable a certain modality usage is for this presentation task. The model demonstrates a possible approach towards cognitive-aware modality planning and can be modified for other applications.

Keywords: Cognitive-aware, multimodal presentation, modality planning, cognitive load, stress, performance, high-load HCI.

1 Introduction

Advanced human-computer interactions are often accomplished through multiple modalities, such as text, images, speech, and sound. Modality planning in HCI is often accomplished in a context-aware manner, i.e. the modalities to be used are selected according to communication goals, user profiles, environmental conditions and resource limitations [1]. However, as multimodal presentations are created for human users to perceive, process and act upon, computer systems should understand not only how to convey information, but also how human minds take in and process the information. When taking into account the modality-related knowledge of human cognition, multimodal presentations could be created in a cognitive-aware manner, so they can be more efficiently perceived and processed. We believe that the cognitive aspects of modality planning are particularly essential in a high-load HCI situation, where the interaction challenges the full capacity of the human cognition.

A huge body of psychology studies provides modality-related cognitive theories and principles that are potentially useful for cognitive-aware modality planning. According to Baddeley's working memory model [4], the working memory has

separated stores (perception channels) for visual information and auditory information, and each store has a limited capacity. Therefore, the capacity of working memory can be better used when both channels are used to perceive information. This theory is known as the dual-channel theory. Another modality-related finding of Baddeley is that the working memory relies on sub-vocal speech to rehearse information and maintain memory traces [3]. Furthermore, the dual-coding theory of Paivio [13] states that verbal and nonverbal information are represented and processed in separated mental systems. These two systems are interconnected through dynamic associative processes. Studies on multimedia learning have demonstrated that the associative processes between the verbal and nonverbal mental systems play a major role in knowledge comprehension and long-term memorization [8, 12].

In this study, we attempt to apply these modality-related cognitive theories as a foundation of cognitive-aware multimodal presentation. An experiment was conducted to investigate the cognitive effects of modalities, using a high-load HCI task. The results showed a significant modality impact on performance, cognitive workload and stress. Based on the experimental findings, we integrate the relevant theories into a prediction model that can systematically compare the suitability of different modality usages for this specific task.

2 Experiment

We created an earthquake rescue scenario, where the locations of wounded and dead people are continuously reported to the crisis response center and displayed on a computer screen. Based on these reports, a crisis manager directs a doctor to reach all wounded people and save their lives. In this experiment, the subject plays the role of the crisis manager and his/her task is to save as many wounded victims as possible.

2.1 Presentation Material

For each victim report, two types of information can be provided: basic information and additional aid. The basic information includes the type of the victim (wounded or dead) and its location. The additional aid reduces the searching area by indicating which half of the screen (left or right) contains this victim.

To convey these two types of information, we selected four modalities based on their visual/auditory and verbal/nonverbal properties: text (visual, verbal), image (visual, nonverbal), speech (auditory, verbal) and sound (auditory, nonverbal). The basic information can be efficiently conveyed by locating a visual object on a map. We use text or image to present a victim (see Fig. 1), and the cell it occupies on a grid-based map indicates the location of the victim (see Fig. 2). The additional aid can be presented by text ('left' or 'right'), image (a left arrow or a right arrow), speech ('left' or 'right') or sound (an ambulance sound coming from the left or the right speaker). Previous studies suggest that the categorization and understanding of concrete objects are slower when they are presented by text than by image [2, 7, 14]. Therefore, in order to better observe the benefit of the additional aid, text is used to present the

basic information if additional aids are given. Finally, five experimental conditions were selected (see Table 1).

Table 1. Five experimental presentation conditions

| Index | Basic Information | Additional aid | Modality properties |
|-------|-------------------|----------------|---------------------------------------|
| 1 | Text | None | Visual, verbal |
| 2 | Image | None | Visual, nonverbal |
| 3 | Text | Image | Visual + visual, verbal + nonverbal |
| 4 | Text | Speech | Visual + auditory, verbal + verbal |
| 5 | Text | Sound | Visual + auditory, verbal + nonverbal |

| | Text | Image |
|---------|---------|---|
| Patient | Patient |  |
| Death | Death |  |

Fig. 1. Text and image presentations of the victim type.

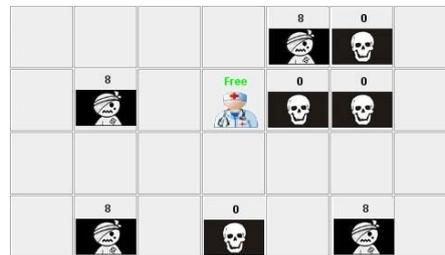


Fig. 2. Sample of the grid-based map (partial)

2.2 Task

The subject played the role of the crisis manager. The task was to send the doctor to each patient by mouse-clicking on the presentation (text or image). A new patient appeared with a random interval between 2 to 5 seconds, usually at the same time as one or more dead victims. A patient had a life time of 10 seconds and would turn into a dead victim without a timely treatment. A number above the presentation of a patient indicated his remaining life time. When timely treated, patients disappeared from the screen. In each trial, 100 patients were presented in about 5 minutes.

The difficulty of the task could be regulated by the number of distracters (dead victims). At the beginning of a trial, there was no object on the grid map and the task was relatively easy. As the number of dead victims grew, it became more and more difficult to identify a patient in the crowded surroundings. The task difficulty reached the maximum (about 40% of the cells contained objects) after about 150 seconds and remained unchanged for the rest of the trial (see Fig. 3).

2.3 Measurements

Three categories of measurements were applied, namely performance, subjective and physiological measurements. The performance in each trial was evaluated by three variables: 1) the average reaction time to click on a patient (in seconds), 2) the total number of patients that were not treated within 10 seconds and 3) the time stamp when the first patient died in a trial (in seconds). The NASA Task Load Index [9] was

used to obtain subjective reports on cognitive workload and stress. A 20-level rating, from very low (1) to very high (20), was performed on cognitive workload and stress, respectively. In order to further assess cognitive workload from the physiological states, we recorded the electrocardiograms, galvanic skin conductance and respiration during the experiment. Scientific literature suggests that when the cognitive demand increases, the heart rate increases, the heart rate variability decreases, the skin conductivity increases and the respiration rate increases [5, 10, 15, 17].

2.4 Subjects and Procedure

20 university students (bachelor, master or Ph.D.) volunteered to participate in this experiment. After entering the lab, the experimenter first applied the physiological sensors to the participant, while he/she was listening to soothing music. When the sensors were set, an additional resting period of 5 minutes was given and then the baseline physiological state was recorded for 5 minutes. Afterwards, the participant received an introduction to the experiment and performed a short training session in order to get familiar with the task and presentation conditions. Finally, the participant performed the five experimental trials with a counterbalanced order. A 5 minutes break was placed between each two successive trials. The subjective ratings on cognitive load and stress were conducted during the breaks. The whole experimental procedure lasted for about 80 minutes.

3 Results

Due to our experimental design, we applied repeated-measure ANOVAs on the dependent measurements, with modality as a within-subject factor. Results from the three categories of measurements are presented in this section.

3.1 Performance measurements

ANOVA results showed a significant modality effect on all the three performance measurements. First, modality had an effect on the average reaction time, $F(2.87, 54.51) = 12.76, p < 0.001$. Subjects reacted the fastest in the 'text + speech aid' condition, spending 1.95 seconds on average to rescue a patient. The average reaction time was the longest (3.05 seconds) in the 'text + no aid' condition. Second, the modality factor also has an effect on the number of dead patients, $F(2.36, 44.84) = 16.81, p < 0.001$. The least patients died in the 'text + speech aid' condition (3 on average), and the most died in the 'text + no aid' condition (11 on average). Third, modality had an effect on the time stamp of the first dead patient in a trial, $F(4, 76) = 17.71, p < 0.001$. The first dead patient occurred the earliest in the 'text + no aid' condition (at the 73th second on average) and the latest in the 'text + speech aid' condition (at the 221th second on average). As mentioned above, the task became more and more difficult as the number of distracters (dead victims) increased.

Therefore, the time difference of the first dead patient indicates that, due to the different modality usages, the performance had different levels of tolerance against the increase of the task difficulty (see Fig. 3).

In general, the ‘text + no aid’ and ‘text + image aid’ conditions form a low performance group; while the other three conditions form a high performance group. Pair-wise comparisons (post-hoc tests) show significant differences between all pairs of conditions that are taken from different groups (see [6] for more details). Moreover, results of these three measurements are strongly positively correlated, indicating that when modalities are more properly used, subjects react faster, fewer patients die and a good performance holds longer.

3.2 Subjective Measurements

The average subjective ratings on cognitive workload and stress are shown in Fig. 4. ANOVA results show a significant modality effect on both the subjective cognitive workload ($F(4, 76) = 16.91, p < 0.001$) and the subjective stress ($F(4, 76) = 9.379, p < 0.001$). There is a strong positive correlation between these two measurements ($Cor. = 0.855$), suggesting that subjects feel more stressed when they devote more cognitive efforts into the task. The experienced stress and cognitive workload are the highest in the ‘text + no aid’ condition and the lowest in the ‘text + speech aid’ condition. Furthermore, the subjective measurements are also positively correlated with the performance measurements, indicating that when the task is more difficult due to an improper usage of modalities, subjects devote more cognitive effort, feel more stressful and perform worse. The subjective measurements also show a two-group pattern. Subjective stress and cognitive workload are significantly higher in the ‘text + no aid’ and the ‘text + image aid’ conditions than in the other three conditions.

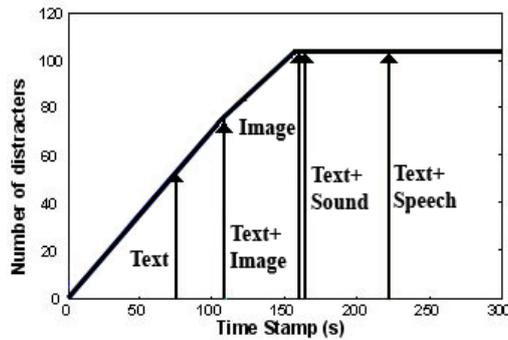


Fig. 3. The time stamps of the first dead patient

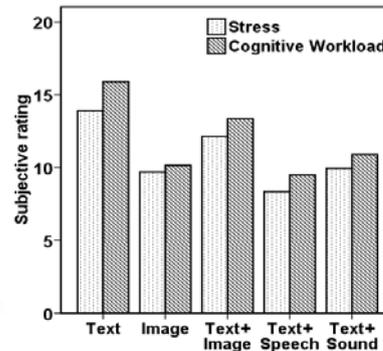


Fig. 4. Subjective rating on cognitive workload and stress

3.3 Physiological Measurements

Table 2 shows the seven physiological measurements that are calculated for each trial (based on [11, 16]). In order to eliminate the individual differences in physiological

activities, the measurements are normalized within each subject, using the baseline values. For example, the average HP value of trial n from subject 1 (s1) is normalized as follows:

$$HR_{n,s1, norm} = \frac{HR_{n,s1} - HR_{baseline,s1}}{\max(HR_{n,s1} | n = 1, 2, \dots, 5) - HR_{baseline,s1}}, \quad n = 1, 2, \dots, 5$$

Table 2. Physiological measurements

| Category | Measurement | Description |
|------------------------|-------------|--|
| Heart period | HP | The time interval between two successive heart beats |
| Heart rate variability | LF | The total spectrum power of a heart period series at band 0.07~0.14Hz (HRV analysis at the frequency domain) |
| | RMSSD | The square root of the mean squared differences of successive heart periods (HRV analysis at the time domain) |
| | NN50 | The number of internal differences of successive heart periods that are greater than 50ms (HRV at time domain) |
| Skin conductivity | GSRN | The number of event-related skin conductance responses |
| | GSL | The tonic level of skin conductivity |
| Respiration | RP | The time interval between two successive respiration peaks |

Using repeated-measure ANOVAs, a modality effect was only found on the LF measurement at the 90% confidence level, $F(4, 76) = 2.52$, $p = 0.07$. This result indicates that the variance in the task difficulty (due to different modality conditions) did not have a significant impact on the subjects' physiological states.

3.4 Discussion

The modality effects on cognitive workload, stress and performance have been consistently confirmed by the performance and the subjective measurements. First, image is a better modality than text to present the basic information. In line with the literature, our results suggest that when presenting concrete objects, the understanding and categorization of images is easier and faster than text. Second, the two 'visual + auditory' modality combinations (see Table 1) significantly outperform the 'visual + visual' combination. Since the basic information already imposes a high load to the visual perception channel, an additional visual aid can only further split up the attention and cause distractions, instead of actually aiding the performance. On the other hand, an auditory aid can be of real help, because it effectively provides extra information without imposing any extra load to the visual channel. This finding is in line with the dual-channel theory of Baddeley [4]. Third, the verbal additional aids significantly outperform the nonverbal additional aids. In this task, it often happens that new patients arrive when one or more earlier presented patients are still unrescued. When additional aids are given, subjects try to maintain a queue of 'left's and 'right's in their minds while searching for the earlier presented patients, because their lifetimes are decreasing as they wait for treatment. Since the working memory relies on verbal codes to maintain memory traces, verbal aids can be directly rehearsed and maintained. However, nonverbal aids require an extra translation through the associative processes between the two mental systems, resulting in a

higher cognitive demand. This finding is consistent with the working memory theory of Baddeley and the dual-coding theory of Paivio [13]. Finally, the ‘text + speech aid’ condition, as a ‘visual + auditory’ combination with verbal aids, is proved to be the best modality condition among the five. Although ‘text + no aid’ is the worst condition, when the additional aid is presented by a proper modality, the combination significantly improves the performance and reduces the cognitive load and stress.

A modality effect on physiological states is not found. Further analyses provide two explanations for this result. First, for each measurement, we applied a t-test, comparing the mean values of the 5 experimental trials to the baseline values. Significant differences were found in all measurements except GSRN. For HP, LF, GSL and RP, the differences were in the expected direction for all 20 subjects. This shows that the physiological measurements did pick up the major changes in cognitive demand between the baseline and the 5 experimental conditions. However, it seems that they were not sensitive enough to reflect the relatively small variances in the cognitive demand between the 5 experimental trials. Second, the level of sensitivity of the physiological measurements might be different for each subject, i.e. a measurement might be sensitive for some subjects, but not all. If so, statistical analyses of the data from 20 subjects would not reveal robust patterns. To prove this explanation, we selected a number of good and bad performance periods from the performance data of each of 5 subjects (randomly selected). Bad performance periods were taken by selecting a 10-second window centered at all time stamps when a patient died. The same number of good performance periods (also 10 seconds long) was taken from the beginning of the five trials when the task was relatively easy. We assume that the cognitive load level was higher in the bad performance periods than in the good performance periods. Six¹ measurements were re-calculated in each period. Then, t-tests were conducted between these two conditions, for each measurement respectively. The results (see Table 3) indeed show individual differences in the sensitivity of physiological measurements. For example, the heart rate measurement (HP) is sensitive for subjects 2, 4 and 5. Heart rate variability (RMSSD, NN50) is sensitive only for subject 3. Skin conductivity (GSL) is sensitive for subjects 2 and 4. Respiration (RP) is sensitive only for subject 5. This finding indicates that the physiological assessment of cognitive workload should take the individual differences between subjects into account, especially when the variances to be detected are relatively minor.

Table 3. Individual differences in the sensitivity of physiological measurements

| Subject index | No. of good p.p. | No. of bad p.p. | Physiological measurement with significant t-test results | |
|---------------|------------------|-----------------|---|------------|
| | | | at 95% cl. | at 90% cl. |
| 1 | 55 | 55 | none | none |
| 2 | 27 | 27 | GSL | HP |
| 3 | 37 | 37 | RMSSD | NN50 |
| 4 | 39 | 39 | HP | GSL |
| 5 | 43 | 43 | RP | HP |

¹ LF is not applicable with a 10s window. Normally, about 300 data points (about 5 minutes) are required to resolve frequencies down to the 0.01Hz level [16].

4 A Suitability Prediction Model

The modality-related cognitive theories are in line with our experimental findings, suggesting that they could be applied as a theoretical foundation for cognitive-aware modality planning. In this section, we demonstrate a possible way of integrating these cognitive theories into a model that can systematically predict the suitability of a certain modality usage for our presentation task. In this task, a modality usage is considered as more suitable if processing the information presented with this modality usage imposes lower cognitive load on the users. Based on the same set of modalities as used in our experiment, the model is able to predict the suitability of all modality combinations, including those that are not investigated in the experiment.

A linear utility function is constructed that takes modalities as inputs and outputs a value describing the suitability level. The higher the output value is, the more suitable the input modality usage is. The function contains three attributes: 1) the representative property of the modality that presents the basic information (B). In our scenario, auditory modalities (speech and sound) cannot be used to present the locations of objects (see [6] for details). Therefore, possible options are text and image. Based on the literature and our experimental findings, image is more suitable than text, thus a 2 is assigned to image and a 1 to text; 2) the perception property of the modality that presents the additional aids (P). Possible options are visual, auditory and none. Based on the dual-channel theory, when a visual modality is used for B, a visual aid causes distraction and harms the performance, while an auditory aid assists the performance. Therefore, a -1 is assigned to the visual modalities and a 1 to the auditory modalities; 3) the mental system the assisting modality belongs to (M). Possible options are verbal, nonverbal and none. According to the working memory theory and the dual-coding theory, verbal aids are more beneficial than nonverbal aids, thus a 2 is assigned to verbal modalities and a 1 to nonverbal modalities.

Furthermore, a weight (f) is assigned to each attribute, determining how much the attribute contributes to the final suitability score. The summary of the three weights is 1. Finally, the suitability prediction model is as follows:

$$\text{Suitability} = f_B \times B + f_P \times P + f_M \times M$$

The choice of weights can be influenced by the task condition. For example, the factor M is less important in a low-load condition than in a high-load condition, because there is less or no pending information to be maintained. The factor P can be very important in an extremely low-load condition (e.g. one patient per hour), because the visual vigilance could be low but the auditory signals have an alerting function. For a high-load situation as in our experiment, the weights are set to 0.5, 0.3 and 0.2 for B, P and M, respectively. The suitability predictions for 10 possible modality usages are shown in Table 4. The outcomes for the five experimental conditions are consistent with the performance results and the subjective rating, indicating the validity of this model. The ‘image + speech aid’ combination is predicted to be the best modality usage for this presentation task in a high-load condition.

This suitability prediction model demonstrates the possibility to quantitatively evaluate the cognitive effects of modalities and systematically select the best modality usage for a specific presentation task. To generalize to other applications, the following aspects need to be considered: 1) the output: how to define suitability based

on the presentation goal; 2) the attributes: what criteria to use to predict suitability based on related theories; 3) the weights: how important is each attribute based on task conditions. Moreover, when contextual aspects need to be taken into account for modality planning, they can be either treated as separate attributes or as conditional switches that selectively determine attribute values or weights. For instance, the B value is currently set to 2 for image and 1 for text. However, for the relatively small group of text-oriented users (1 subject out of 20 in our experiment), the B value should be 1 for image and 2 for text, assuming the user preferences are available.

Table 4. Predicted suitability of 10 possible modality usages

| Index | Modality for basic info. | Modality for additional aid | B 0.5 | P 0.3 | M 0.2 | Suitability score |
|-------|--------------------------|-----------------------------|----------|----------|----------|-------------------|
| 1* | text | none | 1 | 0 | 0 | 0.5 |
| 2 | text | text | 1 | -1 | 2 | 0.6 |
| 3* | text | image | 1 | -1 | 1 | 0.4 |
| 4* | text | speech | 1 | 1 | 2 | 1.2 |
| 5* | text | sound | 1 | 1 | 1 | 1.0 |
| 6* | image | none | 2 | 0 | 0 | 1.0 |
| 7 | image | text | 2 | -1 | 2 | 1.1 |
| 8 | image | image | 2 | -1 | 1 | 0.9 |
| 9 | image | speech | 2 | 1 | 2 | 1.7 |
| 10 | image | sound | 2 | 1 | 1 | 1.5 |

*: experimental conditions

5 Conclusion and Future Work

In this study, we conducted an experiment to investigate the cognitive effects of modality, using a high load HCI task. The performance and subjective measurements consistently show a significant modality effect on cognitive workload, stress and performance, indicating the necessity of conducting modality planning in a cognitive-aware manner. When presenting information for a high-load perception task, such as the one in our experiment, the most suitable modality combination is the one that distributes the information load into two perception channels and provides verbal aids to assist the short-term maintenance of the pending tasks.

Although no relation was found between modality usage and subjects' physiological states, the analyses of physiological data brought an important implication to the physiological assessment of cognitive load. That is to take the individual differences in the sensitivity of physiological features into account, especially when the variances to be detected are relatively minor.

The findings of this experiment are in line with several modality-related cognitive theories that can be applied as a theoretical background for cognitive-aware modality planning. We encoded these theories into a linear prediction model. The model quantitatively predicts the cognitive effects of a modality usage and thus determines how suitable it is for a given presentation task. The components of this model can be re-designed for other applications.

Future work is considered in several aspects. First, the complexity of the presented information can be increased in order to better explore the expressive power of the

text modality. Second, the modality effects observed in this high-load perception task need to be further validated with higher level cognitive tasks, such as reasoning, comprehension and decision making. Finally, the generalization of the modality evaluation model needs further investigations.

Acknowledgments. This research is part of the Interactive Collaborative Information System (ICIS) Project. ICIS is sponsored by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024. We thank C. Mühl and E. L. Abrahamse for their help with setting up the experiment. We also thank the 20 participants for their effort and time.

References

1. Andre, E.: The Generation of Multimedia Presentations. In: Dale, R., Somers, H.L., Moisl, H. (eds.): Handbook of Natural Language Processing. Marcel Dekker, Inc., USA (2000)
2. Bachvarova, Y., van Dijk, B., Nijholt, A.: Towards a Unified Knowledge-Based Approach to Modality Choice. In: Multimodal Output Generation (MOG) pp.7-15. (2007)
3. Baddeley, A.D.: Essentials of Human Memory. Psychology Press, USA (1999)
4. Baddeley, A.D., Hitch, G.J.: Working Memory. The Psychology of Learning and Motivation: Advances in Research and Theory **8**, pp. 47-89 (1974)
5. Boucsein, W., Haarmann, A., Schaefer, F.: Combining Skin Conductance and Heart Rate Variability for Adaptive Automation During Simulated IFR Flight. Lecture Notes in Computer Science **4562**, pp. 639-647 (2007)
6. Cao, Y., Theune, M., Nijholt, A.: Modality Effects on Cognitive Load and Performance in High-Load Information Presentation. In: Intelligent User Interface (IUI), pp.335-344 (2009)
7. Carr, T.H., McCauley, C., Sperber, R.D., Parmelee, C.M.: Words, Pictures, and Priming: On Semantic Activation, Conscious Identification, and the Automaticity of Information Processing. J. Exp. Psychol. Hum. Percept. Perform. **8**, pp. 757-777 (1982)
8. Clark, J.M., Paivio, A.: Dual Coding Theory and Education. Educational Psychology Review **3**, pp. 149-210 (1991)
9. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Human Mental Workload **1**, pp. 139-183 (1988)
10. Kramer, A.F.: Physiological Metrics of Mental Workload: A Review of Recent Progress. In: Damos, D.L. (ed.): Multiple-Task Performance. CRC Press, USA (1991)
11. Malik M: Heart Rate Variability - Standards of Measurement, Physiological Interpretation, and Clinical Use. Circulation **93**, pp. 1043-1065 (1996)
12. Mayer, R.E., Moreno, R.: Nine Ways to Reduce Cognitive Load in Multimedia Learning. Educational Psychologist **38**, pp. 45-52 (2003)
13. Paivio, A.: Mental Representations: A Dual Coding Approach. Oxford University Press, USA (1986)
14. Potter, M.C., Faulconer, B.A.: Time to Understand Pictures and Words. Nature **253**, pp. 437-438 (1975)
15. Scerbo, M.W., Freeman, F.G., Mikulka, P.J., Parasuraman, R., Di Nocero, F.: The Efficacy of Psychophysiological Measures for Implementing Adaptive Technology. TP-2001-211018, NASA Langley Research Center, Hampton (2001)
16. Stern, R.M., Ray, W.J., Quigley, K.S.: Psychophysiological Recording (2nd edition). Oxford University Press, UK (2001)
17. Verwey, W.B., Veltman, H.A.: Detecting Short Periods of Elevated Workload: A Comparison of Nine Workload Assessment Techniques. Applied Experimental Psychology **2**, pp. 270-285 (1996)