International Journal of Occupational Safety and Ergonomics (*JOSE*) 2011, Vol. 17, No. 2, 119–127

Event-Related Assessment of Hypermedia-Based E-Learning Materials With an HRV-Based Method That Considers Individual Differences in Users

Károly Hercegfi

Department of Ergonomics and Psychology, Budapest University of Technology and Economics, Hungary

This paper reports results of an assessment of e-learning materials with the INTERFACE software evaluation methodology. On the one hand, this method of analysis allows us to identify ergonomics problems; on the other, to decide to what extent those problems and their severity concern all users in general and to what extent they depend on the users' type and characteristics. This is the first publication to apply the new marking, export, and statistical features of INTERFACE used in a quantitative analysis of heart rate variability (HRV) curves instead of earlier time-based statistics and qualitative methods. It presents correlations between event-related characteristics of human–computer interaction and the currently required mental effort showed with HRV. The paper also discusses correlations between variables and cognitive-style test scores which indicate the role of individual differences in ergonomics.

human–computer interaction (HCI) software ergonomics
software usability testing and evaluation empirical methods heart rate variability (HRV)
e-learning hypermedia

1. INTRODUCTION

This paper presents a case study in ergonomic assessment of hypermedia-based teaching materials with improved methodology.

The use of computer-supported educational technologies is widespread but requires valid and reliable quality measures. By providing feedback, these quality indicators help teachers in evaluating and comparing different titles, and software developers in improving their work.

The experience gained in relation to this hypermedia-based CD can be transferred to web-based applications. This is possible because of the nature of hypermedia in general; however, this particular CD was developed directly to integrate it with a supporting website (see Hercegfi [1]).

While some systems that test software ergonomics collect various performance measures, very few attempt to measure cognitive effort. The basic advantage of the methodology used in this case study lies in its ability to record continuous online data characterizing the user's current mental effort derived from heart rate variability (HRV) at the same time as other characteristics of human–computer interaction (HCI). In this way a very detailed picture can be obtained; it is a reliable basis for deeper understanding and interpretation of mechanisms underlying HCI. Elementary steps of HCI, like users' different

The author would like to thank students and teachers for their participation; the developers of multimedia materials; Orhidea Edith Kiss Endre, Levente Erdős, Zsuzsanna Hanzlik, and Levente Csányi for their valuable contribution; European Distance and E-learning Network (formerly European Distance Education Network) for the partnership in our Leonardo da Vinci project; Prof. Lajos Izsó and Dr. Eszter Láng for earlier developments.

Correspondence and requests for offprints should be sent to Károly Hercegfi, Department of Ergonomics and Psychology, Budapest University of Technology and Economics, Magyar Tudosok krt. 2 bldg. Q H-1117 Budapest, Hungary. E-mail: <hercegfi@erg.bme.hu>.

mental actions followed by series of keystrokes and mouse-clicks, are the basic and usually critical components of information technology (IT) systems. Measuring mental effort as laid down in international standards on evaluation software [2] is a key aspect of empirical methods. Hence, we need methods that are valid and reliable in monitoring users' mental effort during these elementary steps, which would help identify the weak points in software interface.

To attain this goal, Izsó and his team developed a complex methodology called INTERFACE (INTegrated Evaluation and Research Facilities for Assessing Computer-users' Efficiency). Another paper in this issue of *JOSE* outlines it, especially the HRV channel, and presents validation results [3].

The series of experiments presented in this paper come from the quality assessment of a multimedia learning system "Basics of Information Technology" (in Hungarian)[1]. This author led the team that developed the software; therefore, he was thoroughly familiar with the concepts, the design elements and the expected HCI. After the original project, the scientific analysis of the HCI was continued. At the same time, INTERFACE was further developed. This is the first time the new marking, export, and statistical features of improved INTERFACE are applied. They make quantitative analysis of the HRV curves possible, better than earlier time-based statistics and qualitative methods.

2. METHODS

2.1. Participants

Twenty-four students[2] from two vocational secondary schools participated in the study. They had lessons relevant to the subjects they were studying, in which multimedia materials were used. The students were informed that certain data on them were going to be recorded and were asked to allow placing electrocardiogram (ECG) electrodes on their chest for recording HRV data with the data-collecting module of INTERFACE [3].

The 21 participants of the main series of experiments were 14–16 years old. Thirteen of them studied at a vocational secondary school of economics, 8 studied at a technical vocational school; there were 14 girls and 7 boys. This combination of age and type of school was selected because these materials, in spite of their technical aspects, had been designed for a subject called IT, which is taught at this age in all types of schools.

2.2. Schedule of Sessions

Each session had the following schedule:

1. Collecting data about the current user: (a) filling in a questionnaire with demographic data, academic records, data about the familiarity with the computer and the Internet, etc.; (b) Myers–Briggs type indicator (MBTI) to identify the user's cognitive style. (The psychologists of our department developed the local adaptation of this test.)
2. 2–3-min relaxation followed by a 2–3-min period of mental arithmetic for "calibrating" curves depending on the heart rate variance [3].
3. Free browsing of selected materials for 5 min to become familiar with the style and the controls of the CD.
4. Actual learning task: finding and learning two short subchapters in 10 min followed by taking a 2-min knowledge-acquisition test. This part of the session aimed at having students use the materials in a typical way.
5. Searching task: finding short answers to 11 questions (20–25 min). This was the most important part of the session. The sequence of the questions was to lead the users to situations with various possible usability problems.
6. Interview supported by playback (5 min).

[1] The system was developed within the framework of the Leonardo da Vinci Project "Developing and introducing multimedia teaching materials for vocational education" supported by the European Community (contract HU/97/43022/PI/I.1.1.a/FPI).

[2] Three students participated in a pilot study with the same scenario as the one in the more thorough experiments in which the other 21 subjects participated, too.

This method of analysis allowed us to decide what types of problems were significant for the users, and what types of problems set back the users only slightly. On the other hand, data of 68 independent variables during the first step of this session schedule, and further dozens of variables applicable as independent variables (e.g., the users' behavioural characteristics) derived from the following steps, helped us make a profound analysis of the individual differences in HCI. The high number of independent variables allows us to decide to what extent the problems and their assessed severity concerned all users in general, or how these things depended on the users' type and characteristics.

2.3. INTERFACE Method

After the first step, during all the subsequent ones, the data were recorded with INTERFACE [3]. The system simultaneously investigated the following:

- observable actions and behaviour
  - keystroke and mouse events,
  - current screen content,
  - posture, gestures, and facial expression;
- power spectrum of HRV, i.e., an objective measure of current mental effort.

Although in the literature the term HRV is more frequent, we prefer heart period variability (HPV), where the periods of time between the heart beats are simply the reciprocal values of the heart rate: in practice, the periods of heart beats (the intervals between the consecutive R peaks of the ECG, i.e., RR intervals) can be analyzed more directly, and they can be more expressive. Several studies have shown that an increase in mental effort causes a decrease in the mid-frequency (MF, 0.07–0.15 Hz) power band of the HPV power spectrum [4, 5, 6, 7, 8, 9, 10, 11].

To assess the spectral components of HPV power spectra, Láng and her team developed and successfully used an integrated system for ambulatory cardiorespiratory data acquisition and spectral analysis ISAX [8]. We integrated ISAX into INTERFACE.

The main advantage of the assessment method of the spectral components integrated into our system over previous HPV-based methods [7] is the calculation of the MF spectral profile curve: the MF power of the HPV power spectrum as a quasi-continuous function of time that can indicate changes in mental effort within several seconds as opposed to earlier methods with a resolution of tens of seconds at best.

To analyze only a segment of the time series of the RR intervals, windowing functions are required: in this way, the selected segment (frame) is characterized with the calculated MF power of HPV. When this is done, the frame is shifted further and the spectral analysis is repeated, until the frame reaches the end of the time series. This kind of analysis technique embedded in ISAX is applied by scrolling a constant-size frame in small steps[3]. Thus, the MF power of the HPV power spectrum was automatically calculated for each consecutive frame, and the desired profile curve was obtained.

Important features of the analysis technique embedded in ISAX include the use of the Hamming windowing function and of an all-pole auto-regressive model (instead of, e.g., a spectrum analysis based on Fourier's algorithm) with built-in recursive Akaike's criteria (final prediction error) [12] and a modified Burg's algorithm [13] (for the background of this method and its validation see Hercegfi [3]).

2.4. INTERFACE Method in this Series of Experiments

The variables derived from the INTERFACE records contain time data, number of clicks, number of occurrences of special activities (i.e., the number of times the Back button, Help, or other features of the tested software were used), categorization of recorded tracking data in the hypermedia space, and for the first time in this

[3] In this series of experiments 32-s frames were windowed, and the frame was repeatedly shifted in 2-s steps.

paper, the means of selected parts of the MF profile curve of HPV.

Statistical calculations can use the original absolute MF power of HPV values. However, because of the huge, and natural, individual differences in HPV, the relative changes in HPV are more informative: they show the level of mental effort HCI demands personally of the user.

In this series of experiments, the second step, i.e., the calibration tasks (2–3 min of relaxation followed by 2–3 min of mental arithmetic), was to check the quality of the records of heart rate intervals and the operation of the HRV method. Similar experiments can be used to validate this method [3]. On the other hand, the MF power of HPV results of the relaxation and/or the mental arithmetic phases can be considered a baseline for the evaluation of the further parts of the experiments. However, these tasks (especially relaxation) are strange for the participants: they were just wired and watched by strangers. The perfectness of these tasks depends on individual differences. That is why in the current series of experiments, another approach was used to calculate relative variables: the average of the MF power of HPV values recorded during the third step of the session-schedule was a baseline. This task (free browsing) was familiar to the participants. It can be considered a task that requires low mental effort (not as low as relaxation, of course). Over the variables derived from the original MF power of HPV means of selected parts of the profile curve, new relative variables were calculated this way.

2.5. Statistical Analysis

Some variables can be considered scale measures, others are ordinal or nominal variables only.

- Time data and physiological data (HPV values) are always scale measures. The number of clicks and occurrences are scale, even though sometimes some of these variables have only a few discrete values. MBTI is a validated psychological tool, so its result dimensions can be considered scale, too.
- Some ordinal variables have just a few discrete values (e.g., most variables in demographic questionnaires), while other ones are finely detailed (e.g., means of academic records in a school year); however, if the scales of their measurement are not validated (academic records are not standardized scientifically), they have to be considered ordinal only.
- Gender, type of school, or categories of recorded tracking are nominal variables.

Because of this variety of variables (and the small number of cases, $N_{max} = 21$), the connections between the variables were tested with carefully selected methods.

- Spearman ρ correlation coefficients were used to test connections between scale and ordinal variables.
- Nonparametric Mann–Whitney *U* tests were used to test differences between values grouped by dichotome variables (comparing distributions across groups with grouping variables with two discrete values).
- Nonparametric Kruskal–Wallis tests were used to test differences between values grouped by multiple-value variables (comparing distributions across groups with grouping variable with three or more discrete values).

SPSS version 19 was used.

3. RESULTS

3.1. Participants

A thorough empirical study with 21 subjects and 233 related variables is lengthy, so it was not realistic to increase the number of participants. Naturally, a series of experiments with only 21 participants cannot be fully representative, but they were varied enough:

- previous experience: 9 students had computers for their personal use (not only for family use), 12 students did not; the distribution of the levels of previous computer and Internet experience was close to normal;

- arts or sciences (technical, IT) interests: 10 participants read literature regularly, 11 did not; 7 participants read IT books and/or magazines, 14 did not; as their favourite subjects, 7 students listed IT, 8 students listed maths, 12 students listed arts and humanities subjects (literature, history, etc.); the distribution of maths grades was close to normal;
- general academic records: the distribution of general grade averages of the students was close to normal;
- cognitive style: MBTI showed various types of dominant cognitive functions.

3.2. Calibration Tasks

Figure 1 is a typical INTERFACE result screen of a relaxing user. It shows a typical change in HRV between relaxation and great mental effort.

3.3. Expected Usability Problem

Figure 2 shows a simple, yet frequent problem: the images seem to be hot links, which they are not. This problem was left in the software on purpose to measure its severity. Originally we intended to focus on it.

It is a very simple problem, so no participants complained about it. However, it objectively resulted in unnecessary workload and a substantial waste of time. This type of a problem can be shown to affect all user types.

An analysis of the records showed that 71% of the participants clicked on the images first. They found the real hot links (text rather than images) after an average of 5.3 s of wasted of time. The maximum delay was 80.5 s (Figure 3).

The number of unnecessary clicks and the wasted time caused by this particular usability problem do not correlate with almost any other variables (Spearman ρ values were calculated). Kruskal–Wallis and Mann–Whitney *U* tests did not give significant results. This means that this type of problem is a general one; it affects all types of users.

An analysis of the MF power of HPV values showed the same result. This is so because the time from entering the current page to the first successful click on the hot text correlates strongly with the mean of the relative MF power

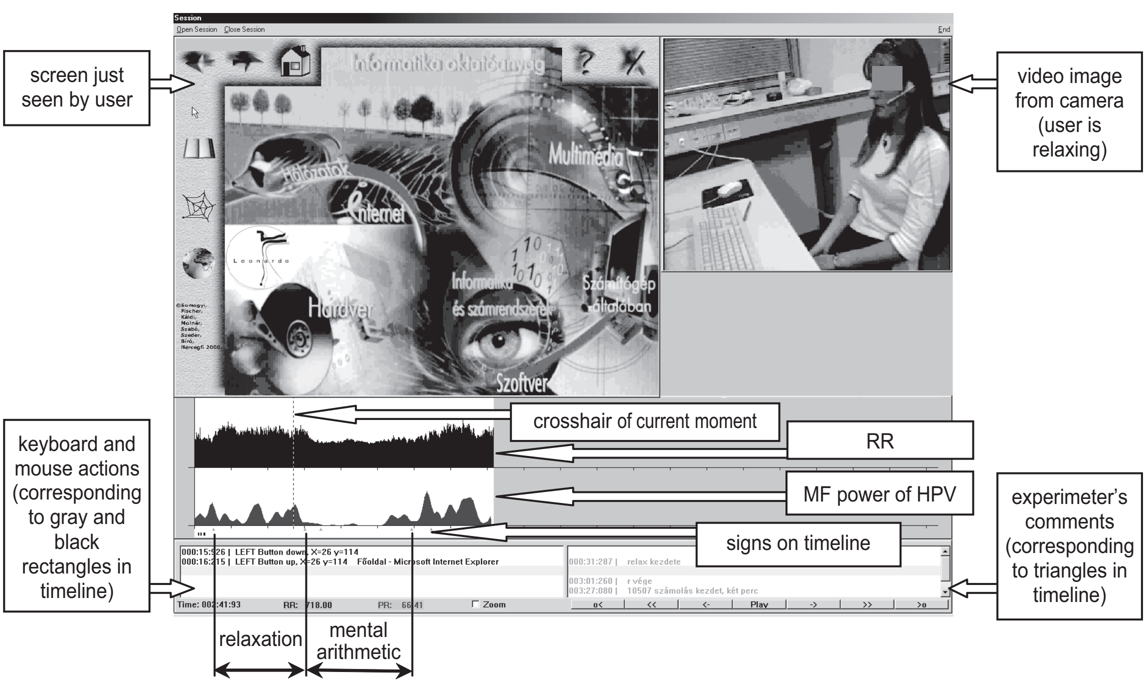

Figure 1. The INTERFACE Viewer screen when the user is relaxing: the profile curve of the mid-frequency (MF) power of heart period variance (HPV) shows relatively high values during relaxation (2 min 30 s) and near-zero values during mental arithmetic (2 min 30 s); the rebound effect is clear immediately after mental arithmetic. *Notes.* RR—periods between heart beats.

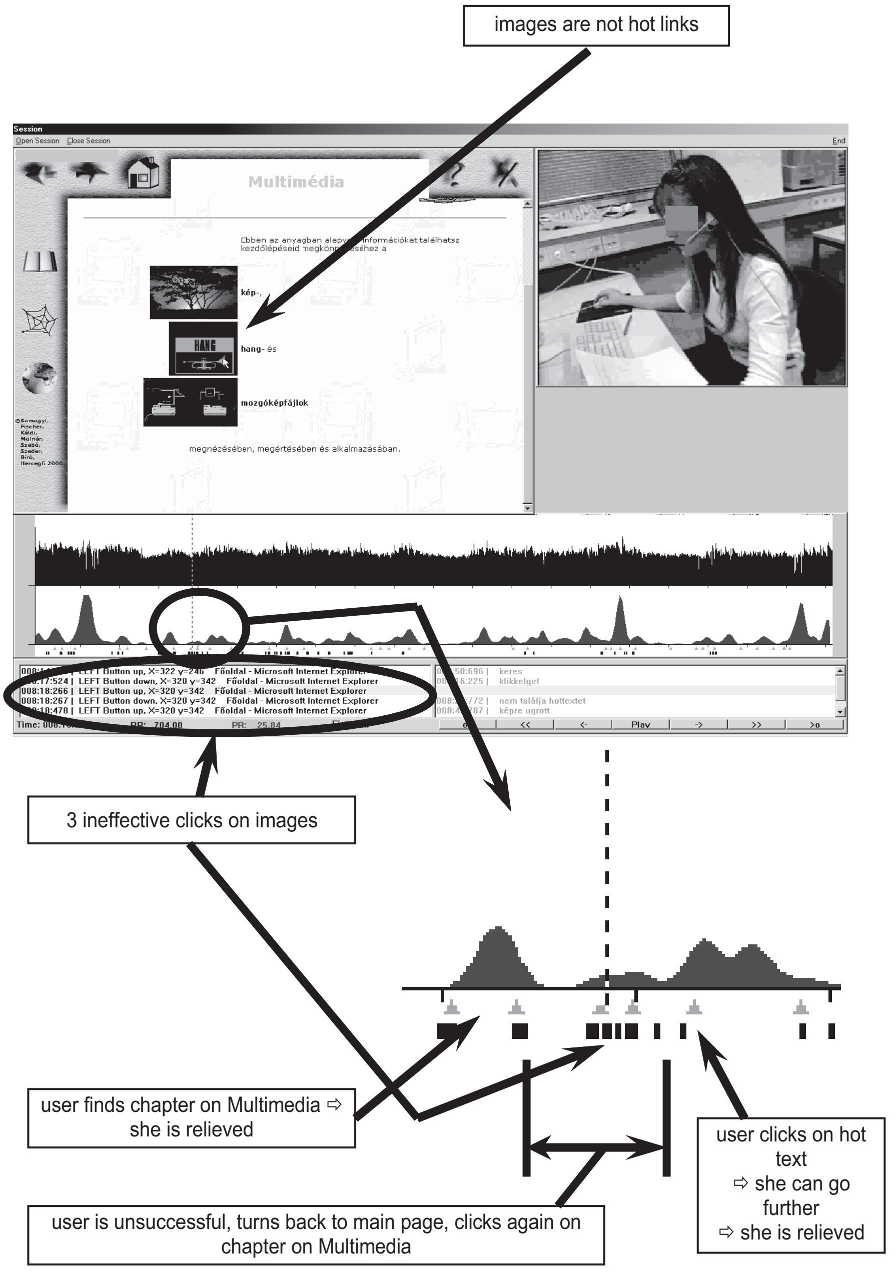

Figure 2. The INTERFACE Viewer screen during a typical software ergonomics problem: the user ineffectively clicks on the images 3 times, which results in a short period of unnecessary mental effort and a loss of 80.5 s.

of HPV during this period: Spearman ρ = –.767 (*p* = .008). This means that usually longer wasted time caused not only longer, but also greater mental effort.

3.4. Unexpected Usability Problem

Figure 3 shows the difficulty in finding the scroll bar. The user discovered the scroll bar only after a 7-min trial-and-error period. The origin of the problem was that the first part of the long, scrollable page, in this screen resolution, looked like a complete page: the figure and its caption were at the bottom of the screen, as seen in the upper screenshot of Figure 3. We did not expect this problem. Its severity depends on the users' type and characteristics.

The average wasted time this problem caused was 69 s. However, it is not just the time that is interesting. One third of the users did not have any problem: they clicked on the scroll bar 1–3 s after they had arrived to this page. The other

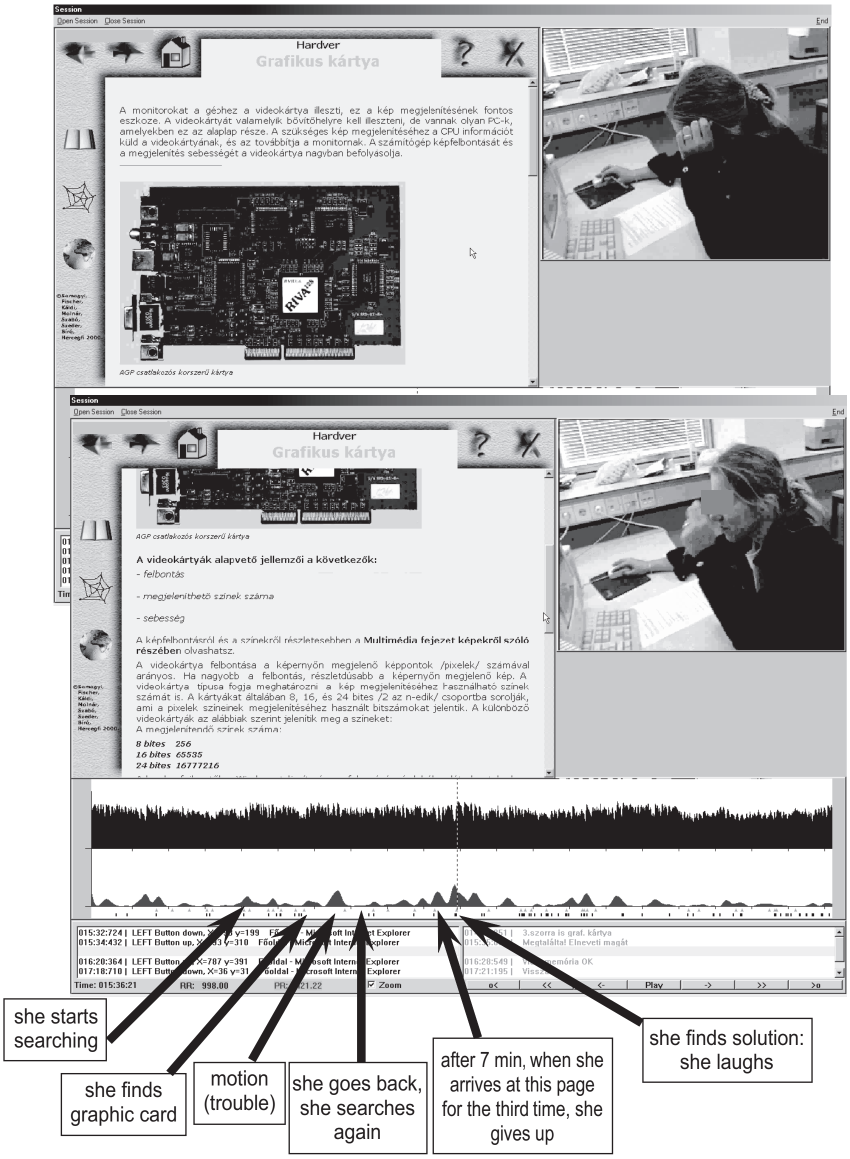

Figure 3. The INTERFACE Viewer screen during a special software ergonomics problem: after 7 min of trying the user gives up, but immediately afterwards she finally discovers the solution (the scroll bar) and laughs. The upper screen shows the moment when the user is giving up, while the lower screen presents the situation when she has just found the scroll bar.

two thirds required 14–253 s. The participant in Figure 3 had no less intellectual capacity or experience with the Internet than the others. Why was this screen a problem for her and the rest of the two thirds of the users, and not for the others? How does the severity of this usability problem depend on the users' type and characteristics?

Mann–Whitney *U* tests showed what time spent on solving the problem depended on.

- Students of the vocational school of economics wasted significantly more time (for type of school, $p = .006$). However, in our case, all the economics students were girls and most of the technical students were boys, so this effect cannot be separated from the effect of the gender (for gender, $p = .031$).
- Users who read literature regularly wasted more time ($p = .021$).
- Users who read IT books and/or magazines regularly wasted less time ($p = .013$).

Calculations of correlations showed that

- students with better grades in mathematics wasted less time (Spearman ρ = –.441, p = .034);
- required time correlated strongly with scores on the thinking–feeling dimension of MBTI (Spearman ρ = –.533, p = .046). So, thinking-type users understood the logic of the content and the user interface almost immediately, independently of the confusing screen, whereas users with a feeling-type cognitive style had a problem with the apparent intactness of the layout of the page.

The quantitative analysis of the MF power of HPV gives us fewer results. In the case in section 3.3., the periods were shorter. There, the participants who needed more time to solve the problem usually made increasingly greater mental effort. However, in the current case, during the long periods the participants sometime gave up for a few seconds, they sometimes did redundant or unnecessary steps without great mental effort, etc. So, the correlation calculations and the other tests did not show any connection between wasted time and the mean of the relative MF power of the HPV recorded during the period. Furthermore, the mean of the relative MF power of the HPV during the period correlated with fewer other variables. However, two important correlations were identified, both are in connection with MBTI.

- The relative MF power of the HPV correlated strongly with scores on the thinking–feeling dimension of MBTI (Spearman ρ = .786, p = .011). This result emphasizes again that thinking-type users understood the logic of the content and the user interface almost immediately, independently of the confusing screen and required less mental effort; but users with the feeling-type cognitive style had a problem with the apparent intactness of the layout of particular page, and it caused more mental effort.
- The relative MF power of the HPV correlated strongly with scores on the extraversion–introversion dimension of MBTI (Spearman ρ = –.874, p = .003). This result means that introvert users were less disturbed by the misleading layout, and required less mental effort.

4. CONCLUSION

The results presented here and in related papers show that in its present form INTERFACE is capable of identifying software ergonomics problems. This methodology can be used to study events in HCI with detail and objectivity other methods do not make possible. The HPV profile function integrated into INTERFACE is a powerful tool for monitoring events in such a narrow time frame that it can practically be considered a time-continuous recording of relevant elementary events.

Beyond the identification of ergonomics problems, in some cases this method of analysis allowed us to decide to what extent the found problems and their assessed severity concerned all the users in general, or how these things depended on the users' type and characteristics.

Even though INTERFACE looks like a quantitative method, its qualitative aspect has to be emphasized. The results in section 3.3. show that sometimes statistical analyses of the relative MF power of HPV give useful results. In other cases, e.g., during longer periods in section 3.4., statistical analyses of HPV provide fewer results than an analysis of simpler time data. The qualitative aspect of the method is more important then: finding mental effort required low parts of the profile curve, understanding the connection between the curves and the events, and recognizing the patterns gave the main results.

REFERENCES

1. Hercegfi K. CD-ROM based multimedia materials integrated with supporting web-sites. In: Proceedings of EDEN (European Distance Education Network) 3rd Open Classroom Conference. Budapest, Hungary: EDEN; 1999.
2. International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC). Software engineer-

ing—product quality (Standard No. ISO/IEC 9126:1991). Geneva, Switzerland: ISO; 1991.

3. Hercegfi K. Improved temporal resolution heart rate variability monitoring—pilot results of non-laboratory experiments targeting future assessment of human–computer interaction. International Journal of Occupational Safety and Ergonomics (JOSE). 2011;17(2):105–17.
4. Orsila R, Virtanen M, Luukkaala T, Tarvainen M, Karjalainen P, Viik J, et al. Perceived mental stress and reactions in heart rate variability—a pilot study among employees of an electronics company. International Journal of Occupational Safety and Ergonomics (JOSE). 2008;14(3):275–83.
5. Mulder G, Mulder-Hajonides van der Meulen WREH. Mental load and the measurement of heart rate variability. Ergonomics. 1973;16:69–83.
6. Rowe DW, Sibert J, Irwin D. Heart rate variability: indicator of user state as an aid to human-computer interaction. In: Proceedings of CHI98—Conference on Human Factors in Computing Systems. New York, NY, USA: ACM Press; 1998. p. 18–23.
7. Mulder G, Mulder LJM, Meijman TF, Veldman JBP, Roon AM. A psychophysiological approach to working conditions. In: Backs RW, Boucsein W, editors. Engineering psychophysiology: issues and applications. Mahwah, NJ, USA: Erlbaum; 2000. p. 79–110.
8. Izsó L, Láng E. Heart period variability as mental effort monitor in human computer interaction. Behav Inf Technol. 2000;19(4):297–306.
9. Izsó L. Developing evaluation methodologies for human-computer interaction. Delft, The Netherlands: Delft University Press; 2001.
10. Wilson GF. Psychophysiological test methods and procedures. In: Charlton SG, O'Brien TG, editors. Handbook of human factors testing and evaluation. 2nd ed. Mahwah, NJ, USA: Erlbaum; 2002. p. 127–56.
11. Lin T, Imamiya A. Evaluating usability based on multimodal information: an empirical study. In: Proceedings of ICMI '06—the 8th International Conference on Multimodal Interfaces. New York, NY, USA: ACM Press; 2006. p. 364–71.
12. Akaike H. Fitting autoregressive models for prediction. Ann Inst Statist Math. 1969;21:243–7.
13. Gray AH, Wong DY. The Burg algorithm for LPC speech analysis synthesis. IEEE Trans Acoust Speech Signal Process. 1980;28(6):609–15.