International Journal of Occupational Safety and Ergonomics (*JOSE*) 2011, Vol. 17, No. 2, 105–117

Improved Temporal Resolution Heart Rate Variability Monitoring—Pilot Results of Non-Laboratory Experiments Targeting Future Assessment of Human–Computer Interaction

Károly Hercegfi

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

This paper outlines the INTERFACE software ergonomic evaluation methodology and presents new validation results. The INTERFACE methodology is based on a simultaneous assessment of heart rate variability, skin conductance, and other data. The results of using this methodology on-site, in a non-laboratory environment indicate that it is potentially capable of identifying quality attributes of elements of software with a temporal resolution of only a few seconds. This paper presents pilot results supporting this hypothesis, showing empirical evidence in spite of the definitely non-laboratory environment: they indicate that the method is robust enough for practical usability tests. Naturally, in the future these pilot results will have to be followed with further laboratory-based verification and refinement. This paper focuses only on some characteristics of this method, not on an actual analysis of human–computer interaction; however, its results can establish a future practical and objective event-related analysis of software use.

human–computer interaction (HCI) software ergonomics software usability testing and evaluation empirical methods heart rate variability (HRV) skin conductance

1. INTRODUCTION

Usability is the key word in assessing the ergonomic quality of software products. The mental effort required by current human–computer interaction (HCI) is an important usability factor. In some definitions of usability as a quality dimension of software products, as laid down, e.g., in the original[1] version of Standard No. ISO/IEC 9126:1991 [1], the required mental effort appears not only as a metric, it is the core of the definition: the better its quality, the lesser the effort required when using the software.

Evidently, in addition to subjective methods, such as questionnaires, objective methods are also necessary in measuring mental effort (self-imposed mental work stress). This paper focuses on heart rate variability (HRV) power spectrum as a technique for measuring current mental effort as a function of time. It also presents a brief overview

The author would like to thank Prof. Lajos Izsó and Prof. Eszter Láng for the earlier developments, our research fellows and industrial partners for the support of the research, and the participants of the series of experiments for their valuable contribution.

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>.

[1] Unfortunately, the definition of usability was drastically simplified in the later version of this standard (ISO/IEC 9126-1:2001) only to a satisfaction-like term "The capability of the software product to be understood, learned and attractive to the user…". We consider this change to be a wrong direction. Standard No. ISO/IEC 9126-1:2001 has been withdrawn to be replaced with ISO/IEC 25010:2011 published in March 2011, after this paper was already written. It introduces a much more complex model to describe the quality-in-use phenomenon, defining a series of quality characteristics with no general definition of usability, however.

of other physiological channels used in studying HCI. Those channels will be listed as they appear in the literature on ergonomics: this paper does not intend to analyze their validity or practical diagnostic value. However, this paper focuses on some characteristics of a special HRV-based method used in the INTERFACE methodology developed at the Budapest University of Technology and Economics.

This paper considers only some characteristics of this method, it does not present any particular usability problems yet; however, its results can establish future practical and objective event-related analysis of HCI.

1.1. Assessing Mental Effort via Analyzing Users' HRV Power Spectrum

Sometimes heart rate itself is used in usability evaluations; however, it is not a sensitive measure of mental effort and thus usability.

The deviation (or variance) of the user's heart rate can give us much better results, but the sources of the variability also include physiological mechanisms independent from the mental effort. Because of this, further spectral analysis of HRV is needed. Although in the literature the term HRV is more frequently mentioned, we prefer a similar expression, heart period variability (HPV), where the periods of time between the consecutive heart beats are simply the reciprocal values of the heart rates. In practice, the periods of heart beats can be analyzed more directly, and they can be more expressive.

The periods between heart beats are called RR intervals, because they are determined by the highest peaks (R peaks) of the electrocardiogram (ECG) curve. (In spite of this, in some papers the RR intervals are referred to as the easy-to-remember rythm-to-rythm intervals [2].)

After an analysis of the variability of the RR intervals, several studies showed that an increase in mental load caused a decrease in the mid-frequency (MF, 0.07–0.15 Hz) power band of the HPV power spectrum [2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. Focusing on this frequency band filters other peaks of the power spectrum: a typical peak in the 0.15–0.45 Hz band corresponds to the respiratory rate (called respiratory sinus arrhythmia); a peak in the 0.04–0.07 Hz band is connected with the thermoregulatory fluctuations of the organization [5, 8]. Heart rate fluctuations in the MF power band may reflect postural changes (via the blood pressure control of the baroreflex). To separate the effect of the mental load from the effect of postural changes, a ratio of the MF component of ~0.1 Hz and a higher frequency respiratory component can be used [4]. However, if the participants work continuously in a sitting posture (e.g., when using a computer), and their larger muscle movements (stretching, laughing, sneezing, talking, etc.) eventually are filtered from the records (e.g., via video analysis), the MF power band itself can characterize the mental effort sensitively enough, as it is shown by the following results presented in this paper.

To assess the spectral components of HPV power spectra, Láng and her team developed and successfully used an integrated system called ISAX (integrated system for ambulatory cardio-respiratory data acquisition and spectral analysis) [5, 8, 12, 13]. We integrated this equipment and the method into our INTERFACE system.

Decisions in the cortex can be made in tenths of seconds. The vegetative control mechanisms of the brainstem are faster. The speed of the impulses (conduction velocities) in the preganglionic fibres of the autonomic nervous system ranges between 3 and 15 m/s, while the speed in the postganglionic fibres is 0.25–1.5 m/s [14]. Therefore, calculating with the slowest signal propagation and the longest distance between the brain and the heart, the neural impulse must arrive to the sinus node from the brain in under 0.5 s. This means that alteration in the HPV power spectrum may occur in a few seconds.

These considerations have driven us to attempt to study these opportunities in a practice-oriented pilot empirical study. What we need for practical purposes is the MF power of the HPV power spectrum as a quasi-continuous function of time. If we had such a curve with good-enough temporal resolution, it would be possible to systematically attribute certain salient parts of this curve to observed events of HCI.

To achieve this goal, first let us analyze only one segment at the very beginning of the time series of the RR intervals. This can be performed with windowing functions; in this way a selected segment (frame) can be characterized by the calculated MF power of the HPV. When this has been 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 is embedded in the ISAX system: this windowing technique is applied by scrolling a constant-size frame in small steps (in this series of experiments 32-s frames were windowed, and the frame was repeatedly shifted in 1-s steps). In this way the MF power of the HPV power spectrum was automatically calculated for each of the consecutive frames, and it resulted in the desired quasi-continuous time curve, the MF spectral profile curve. As the maximal delay of detection equals the step size, the change in HPV in principle can be reflected in the spectral profile curve within one second. (These spectral profile curves are the lowest curves in Figures 3–5.)

The main advantage of our method over previous HPV-based methods [7] is that the MF component of HPV can indicate changes in mental effort within several seconds as opposed to the earlier methods with a resolution of tens of seconds at best. This feature was achieved with

- an appropriate windowing data processing technique with the Hamming windowing function to decrease the aperture effect and improve the spectrum image;
- the spectral profile curve created on the basis of overlapping windows, by finding the best compromise between spectral and temporal resolution;
- an all-pole auto-regressive model (instead of, e.g., spectrum analysis based on Fourier's algorithm) with built-in recursive Akaike's criteria (final prediction error) [15, 16] and a modified Burg's algorithm [17].

This paper does not intend to analytically prove the possibilities of those algorithms; however, it attempts to study their practical applicability empirically.

1.2. Other Physiological Channels Applied to Studying HCI

There exist several other physiology-based techniques to analyze HCI. Some aim to measure actual mental effort, while others aim to identify emotional aspects of HCI. Emotions can represent a dimension independent of mental effort; however, its importance can also be equally high.

Changes in the electrical activity of the skin (electrodermal activity) can be evoked with various physical and emotional stimuli. We use parameters derived from skin conductance (SC) responses, especially the alternating current (AC) component of SC.

Although there are other techniques for measuring mental effort and emotions, either they are more difficult to evaluate and more invasive (e.g., the electroencephalograph, EEG), or they give an overall, averaged indicator for a relatively long period, from minutes to hours (e.g., visual critical flicker frequency (CFF) and practical applications of biochemical measures).

EEG requires more electrodes than ECG, its electrodes have to be positioned more carefully, and the participants experience it as more invasive. Furthermore, it results in much more complex curves. Various effects have to be filtered from the data, such as the effects of eye blinks [10, 18, 19]. Thus, ECG is a simpler and more preferred method in objectively identifying mental effort.

However, using EEG can be a potential direction of further developments of our methodology: not to simply identify mental effort, but (a) to identify more complex mental or emotional state patterns (using complex methods to analyze complex curves [20, 21]), or (b) to attempt to localize active brain regions (with over 20 [10] or 128- or 256-channel dense array EEG (dEEG) [22, 23]).

Electromyography (EMG) measures muscle activity by detecting surface voltages that occur when a muscle is contracted. In isometric conditions (no movement) EMG is closely correlated with muscle tension. When used on the jaw, EMG is a very good indicator of tension in an individual due to jaw clenching. EMG has been used to distinguish between positive and negative emotions on the face. EMG activity over the brow region (the frown muscle) is lower and EMG activity over the cheek (the smile muscle) is higher when emotions are mildly positive, as opposed to mildly negative [24]. Because of the small sizes (the distance between the electrodes is only ~5 mm) and the closeness of the muscles of the different mimic functions, the electrodes have to be positioned extremely carefully [25]. Furthermore, participants experience electrodes on the face or head as more invasive than those on their fingers measuring SC. Thus, SC is used as the simpler and preferred method in identifying emotional reactions instead of EMG's potential capability of differentiating positive and negative emotions.

Members of our team have also measured mental effort with visual CFF and with biochemical measures, e.g., the level of cortisol in the saliva [9]. However, these methods give only an overall, averaged indicator for a relatively long period of time, from minutes to hours; this is not the temporal resolution that our INTERFACE methodology targets.

Eye-tracking is a promising direction of further developments of our methodology: (a) it is reliably capable of localizing the user interface elements that cause high mental effort or emotional reactions identified by the other physiological channels by synchronizing the channels, and (b) it can be analyzed deeper, deriving parameters referring to the state of the nervous system [26].

Pupillometry (measuring the current diameter of the pupil) is a measurement option that is often accomplished with eye-tracker equipment. It reflects both the users' mental effort and their emotions [18, 27, 28]. It can validate the other physiological channels of our methodology. Eye-tracking and pupillometry are used in our ongoing INTERFACE research [29, 30].

2. METHODS

2.1. INTERFACE Methodology

Izsó and his team developed a complex methodology at the Budapest University of Technology and Economics [8, 9, 31, 32, 33, 34]. Figure 1 shows the conceptual arrangement of an INTERFACE (INTegrated Evaluation and Research Facilities for Assessing Computer-users' Efficiency) workstation. The advantage of this methodology lies in its capability of recording continuous on-line data characterizing the user's current mental effort derived from HPV and the user's emotional state indicated with SC parameters synchronized with other characteristics of HCI, such as screen captures and a log of all mouse and keyboard use input. After careful consideration, the detailed picture thus obtained can be a basis for a deeper understanding and interpretation of psychological mechanisms underlying HCI.

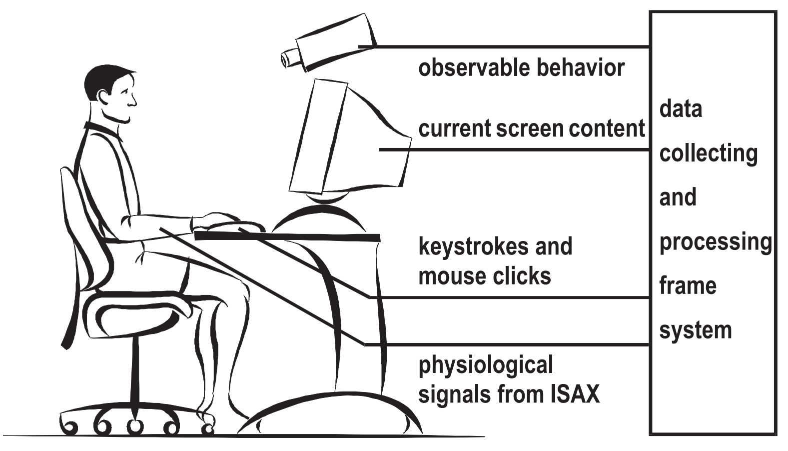

Figure 1. Conceptual arrangement of the INTERFACE software ergonomic testing workstation. *Notes.* ISAX—integrated system for ambulatory cardio-respiratory data acquisition and spectral analysis [5, 8, 12, 13].

At the same time the INTERFACE investigates

- users' observable actions and behavior:
  - keystroke and mouse events;
  - current screen content;
  - users' facial expression, and posture and gestures;
- psychophysiological parameters:
  - power spectrum of HPV, an objective measure of current mental effort [8, 9, 32, 33, 34];
  - SC parameters, indicators of emotional reactions [34].

In contrast with our earlier experiments with HPV only, which we conducted for over 15 years, integrating SC into the INTERFACE methodology is relatively new. Laufer and Németh analyzed SC responses in a series of experiments with a new version of ISAX [35]. This is a good example of using data mining techniques in empirical usability studies. In their studies, the SC measurement was not yet integrated into the INTERFACE system.

In addition to observable elements of behavior, the applied complex method also includes traditional interviews to assess mental models, subjective feelings, and the users' impressions about perceived task difficulty and experienced fatigue.

Recording these various data simultaneously requires more technical resources than other empirical methods based on personal observation or a simple video recording only. However, the synchronization among multiple channels enables researchers to accurately identify and to attempt to interpret significant events during the interaction.

2.2. Experimental Arrangement and Participants

Figure 2 shows a typical experimental arrangement of the INTERFACE methodology as it was used in a recent usability study of

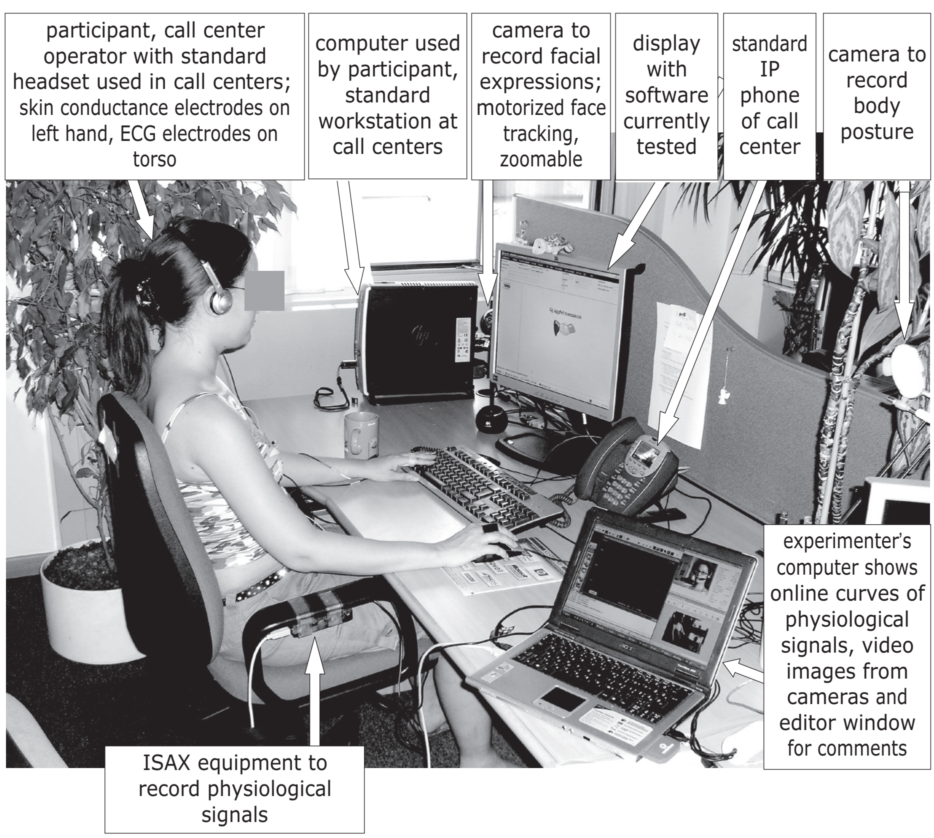

Figure 2. The experimental arrangement during sessions of the INTERFACE usability test, installed on a standard workstation of a call center. *Notes.* ECG—electrocardiogram, IP—Internet protocol, ISAX—integrated system for ambulatory cardio-respiratory data acquisition and spectral analysis [5, 8, 12, 13].

the customer service software of Generali-Providencia Insurance Company (Hungary) [31]. In this study, the user sessions took place at the company's actual call center. The workstation was located in the corner of the operators' room, not to disturb others. The experimenter sat next to the participant. The team leader's glass partition was located behind them, so our staff could sit and observe the sessions and make simulated phone calls from behind this pane.

Three ECG electrodes were placed on the user's torso and two electrodes were placed on their left hand (for right-handed persons) to measure SC. The arrangement of the video cameras and the other equipment can also be seen in Figure 2.

Twelve real operators of the customer service call center participated in the study. In accordance with the typical gender-ratio in this call center job, 10 of them were female, 2 were male. Their age varied (25–55 years, *SD* 9.7). All of them had at least one-year work experience in their current job using the tested software.

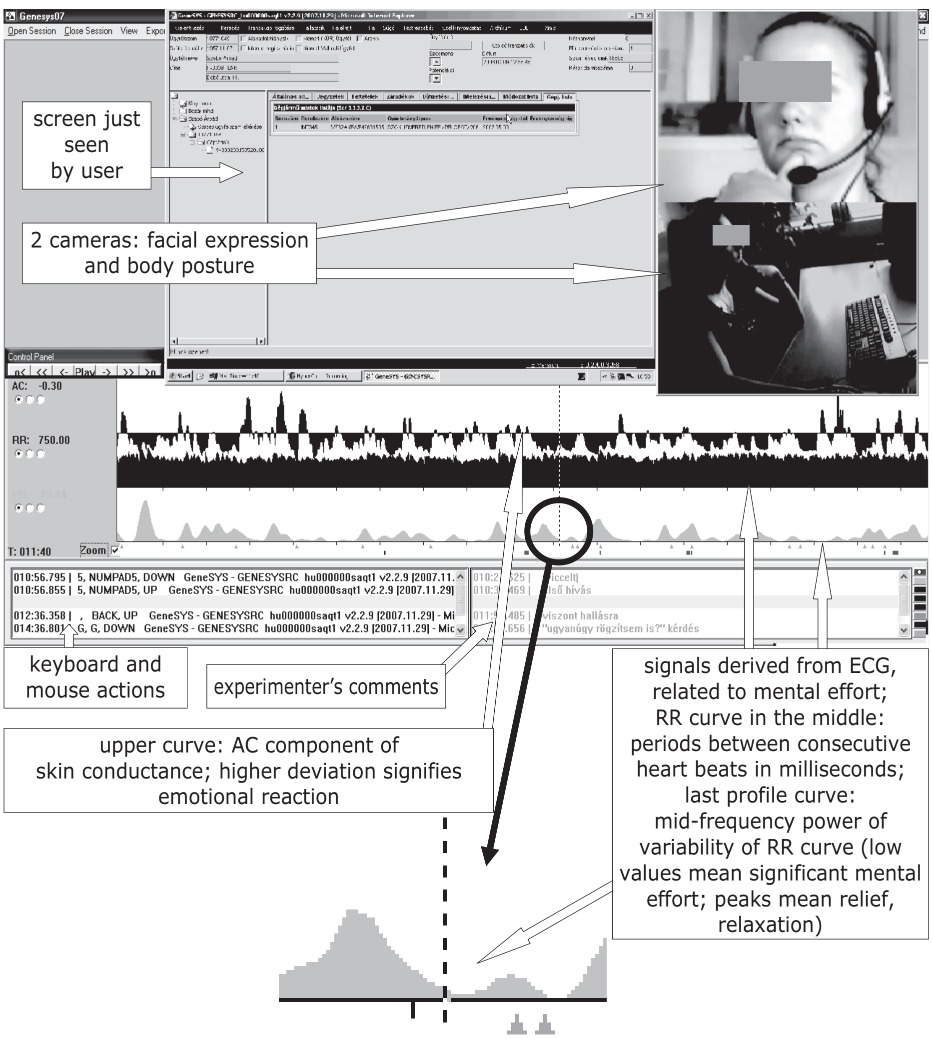

Figure 3. The INTERFACE Viewer screen with a record of the empirical test of call center software. The user at the sample point is exerting significant mental effort (cf. user's facial expression and gesture, and the low value of the last profile curve of the mid-frequency power of the heart period variability at the crosshair). ***Notes*****.** The curves in the window show the history of 20 min: 11 min before the crosshair plus 9 min after the crosshair. The enlarged valley of the profile curve shows a period of 38 s. It is a robust high mental effort period, selected as an illustration; however, the much smaller valleys and peaks can be analyzed and interpreted as well; AC—alternating current, ECG—electrocardiogram, RR—periods between heart beats.

2.3. The Viewer Screen of the INTERFACE Software

The most important strength of the INTERFACE Viewer software is its ability to synchronize and play the records of the different data channels strictly simultaneously. Figure 3 shows the INTERFACE Viewer screen with a record of the empirical test of the call center software described in section 2.2. It also shows a typical pattern of mental effort observable both on the HPV curve and in the video images.

2.4. Meaning of the HPV Profile Curve: Preliminary Impression

Our team at the Department of Ergonomics and Psychology applied the INTERFACE methodology to various evaluation studies [8, 9, 32, 33, 34]. Figure 4 shows examples of the diversity of the software assessed. The case studies in that figure represent assessing (a) ArchiCAD, leading architectural computer-aided design (CAD) software released by Graphisoft; (b) the air traffic control system of the European Organisation for the Safety of Air Navigation (Eurocontrol); (c) WAP-based software from Nokia; and (d) the web-based editor interface of the Moodle (moodle.org) learning management system.

A sample moment of each recording shows a similar situation to the one in Figure 3: in each case the user is exerting significant mental effort, as it is shown by the facial expression, gesture, and posture, and the low value of the last profile curve of the MF power of HPV at the crosshair.

Figure 4. Examples of various types of software assessed with the INTERFACE methodology. The sample moment of each record shows a similar situation to the one in Figure 3: in each case, the user is exerting significant mental effort (cf. user's facial expression, gesture, and posture, and the low value of the last profile curve of the mid-frequency power of the heart period variability at the crosshair). *Notes.* Eurocontrol—European Organisation for the Safety of Air Navigation, WAP—wireless access protocol.

2.5. Validation Methods of the Series of Experiments

This paper presents the main results from recent validation studies with the new statistical features of the INTERFACE frame software as well as our earlier results [8, 9].

At the beginning of each session of our series of experiments, there was a calibration phase. First, the participant was asked to relax for ~2 min. This was followed by a 2-min high mental effort exercise: mental arithmetic.

The instructions for the relaxation periods were always the same; they involved

- the participants seating themselves in a comfortable posture, without any movement;
- the participants keeping their eyes open;
- the participants trying to think about nothing (we knew this was difficult for untrained people; we expected at least trying to avoid to think about specific items);
- calming the participants and assuring them that there were no good or bad personal results: we had no expectations, we only wanted to study some differences between this period and the next one.

The instructions of the mental arithmetic periods were also always the same:

- no movements;
- no speech, no counting aloud, no voiceless movement of the mouth;
- after the experimenter gave the participants a starting number (e.g., 11 789), they were immediately to count backwards for 7 s.

Two minutes later, the result of the counting was asked for; however, the actual result of the counting was not important; the only goal was to artificially generate mental effort.

This usability study was carried out on-site at the insurance company's call center [34]; Figures 2–3 show it did not resemble a laboratory environment. However, we aimed to create, apply, and test a robust enough methodology. In this series of experiments, because of the real-life experimental situation of a working call center, the participants were continuously disturbed by their colleagues' calls and discussions. If the method is robust enough, the differences between the used metrics (the MF power of HPV) corresponding to the periods are statistically significant. (Naturally, because of these circumstances, the possible positive results can be considered only pilot results, and future laboratory verification is still required.)

Twelve operators were involved as participants, each had a one-hour session that was recorded. The quantity of data gained from these sessions was really significant considering the depth of the enquiry. This paper, however, focuses only on the calibration phases of these sessions. These calibration periods make validating the method possible.

First, the validation of the meaning of the curves is targeted. Focusing on the HPV profile curve, the differences of the MF power of HPV values during the relaxation phase, the mental arithmetic phase and the following software use phase can be analyzed.

Next, the validation of the temporal resolution of the HPV profile curve is targeted. As mentioned, the second part of the calibration phase was a 2-min high mental effort exercise (mental arithmetic). The experimenter gave the participants a starting number (e.g., 11 789), and the participants immediately had to count backwards for 7 s. The exact end of the experimenter's instructions can be considered the starting point of the mental arithmetic exercise. On this basis, the speed of the change of the curve can be validated: the differences between the MF values of HPV one second after the start of the exercise and the MF values of HPV one second before the start of the exercise can be analyzed.

2.6. HRV Analysis, Statistical Analysis

The ECG peaks were recorded with ISAX (see section 1.1.). The collected raw ECG peaks and the power spectrum analysis were processed with ISAX software. The following parameters were used to create the spectral profile curve:

- MF band;
- size of the windowing frame: 32 s;

- steps (shifts) of the windowing frame: 1 s.

Because of the low number of participants, the normality of distribution could not be proved. Therefore, a non-parametric statistical method, the Wilcoxon signed ranks test was used to test the differences. SPSS for Windows version 16.0 was used for statistical analysis.

3. RESULTS

3.1. Results of Validation of the Meaning of the Curves

The statistical analysis of the records of the 12 sessions shows validity even in this real-life experimental situation, where the participants were continuously disturbed by their colleagues' calls and discussions. The curves in Figure 5 were recorded during two sessions (No. 10 and No. 11) of the series of experiments carried out in the call center of the insurance company, during relaxation and mental arithmetic exercises.

The upper curves represent AC of SC, the middle ones show the RR values (heart periods), and the bottom curves display the MF power of HPV. The curve of AC of SC is relatively smooth during both relaxation and mental arithmetic. During these sections, there are not any emotional peaks, and these 2 participants can be characterized as stable, according to the typology of physiology. However, peaks follow the beginnings and the ends of the sections.

During relaxation, the MF component of the HPV increases, so the RR curve has zigzags, and the profile curve is relatively high. (During perfect relaxation, the profile curve should be consistently high. However, this was not expected in this experimental situation: this was a real-life situation, the participants were continuously disturbed, and they were wired up and observed.) The curve can be considered high, especially in comparison with the next section.

During the mental arithmetic exercise, the RR curve becomes smoother, and consequently the profile curve is much lower.

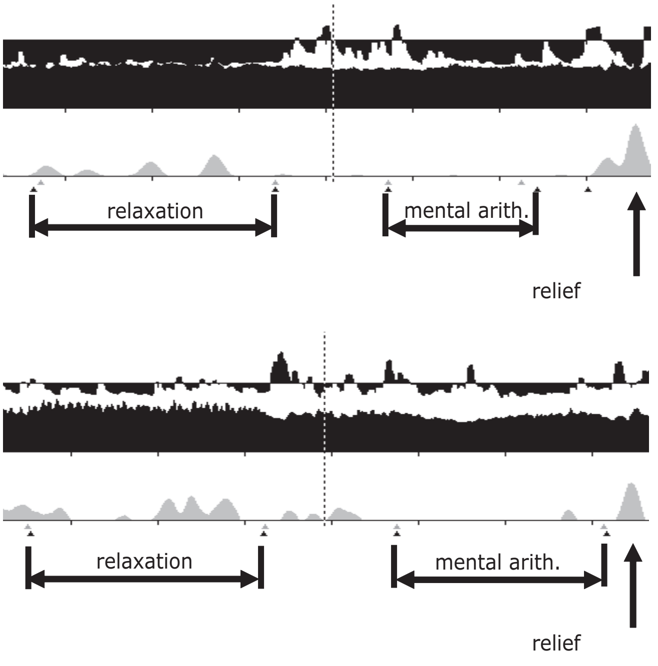

Figure 5. A typical pattern of relaxation and mental arithmetic periods in 2 participants. The upper curves represent the alternating current of skin conductance, the middle curves show the RR values, and the bottom ones display the mid-frequency power of heart period variability. *Notes.* In these cases the relaxation periods lasted 2 min 53 s and 2 min 41 s, the mental arithmetic periods lasted 1 min 42 s and 2 min 24 s; arith.—arithmetic; RR—periods between heart beats.

After the calibration tasks, the participants are relieved. During this short period of relief, the participants are more relaxed than during the conscious, intended relaxation: the MF of HPV profile curves have their highest peaks (the rebound phenomenon).

The values of the MF power of the HPV were significantly higher during relaxation than during mental arithmetic. The Wilcoxon signed ranks test proved the difference (p = .037, Figure 6). This is a significant difference between the MF power values corresponding to relaxation and to mental arithmetic, in spite of the non-perfect relaxation and the real-life working environment.

However, the mental arithmetic task works even better: the significance of the difference between the values of MF power of HPV during mental arithmetic and in general, during the whole software use section is stronger (p = .002).

As expected, there was no significant difference between relaxation and mental arithmetic in the values of the deviation of the AC component of SC. However, the deviations were significantly lower than in general, during the whole software use section (p = .009 and p = .017, respectively).

These results show that low values of the MF power of HPV indicate mental effort, and high deviations of AC of SC probably indicate higher emotions.

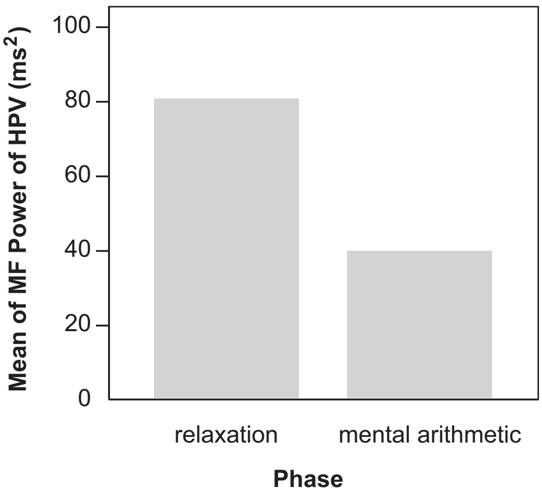

Figure 6. Validation of measuring mid-frequency (MF) power of heart period variability (HPV) as an indicator of mental effort: the MF power of HPV was significantly higher during relaxation than during the mental arithmetic exercise (p = .037).

3.2. Results of Validation of the Temporal Resolution of the HPV Profile Curve

Starting a mental arithmetic exercise causes quick changes in the MF power of HPV. This validates the temporal resolution of the method. The MF of HPV profile curve of the INTERFACE shows this change with a 2-s temporal resolution: the MF of HPV values 1 s after the start of the mental arithmetic exercise are significantly lower than the MF of HPV values 1 s before its start. The Wilcoxon signed ranks test proves the difference (p = .028).

There is a significant difference within this 2-s period, even though the seconds immediately before the starting point of the exercise should not be described as a relaxation period, but attention-requiring listening for the starting number. However, as it can be seen, listening for the starting number requires significantly less effort than the exercise itself.

It is important to emphasize that during these calibration tasks there were 2-min relaxation and 2-min mental arithmetic periods; however, we were able to detect statistically significant differences within that 2-s period around the "changing point". This result is promising, it shows that it might be possible to study in this way short events in future experiments. Future laboratory-based series of experiments focusing on differences in short controlled events are also necessary.

4. DISCUSSION

After validating the series of experiments, in the sections on software use, we looked for moments with relatively high (unwanted) mental effort and high (unwanted and negative) emotions. This method is the key to finding problems in the user interface.

The results show that already in its present form the INTERFACE methodology is capable of identifying relatively weak points in HCI. This methodology made it possible to study events in HCI in temporal resolution and objectivity other currently available methods do not provide. The applied HPV profile function integrated into

INTERFACE is a potentially powerful tool for monitoring events in such a narrow time frame that is practically a time-continuous recording of relevant elementary events.

Naturally, it is necessary to analyze synchronized records of physiological data and videos. With video recordings, e.g., the artefacts of HPV caused by large movements (such as stretching) can be filtered out: these peaks in the profile curve cannot be interpreted mechanically as a decrease in mental effort. In other cases, peaks in the profile curve can indicate relaxed periods during easy use of software or they can indicate giving up (not coping any more): they can be differentiated on the basis of the participant's efficiency. (Even if there are objective performance metrics, however, this interpretation has to be careful.)

After artefact filtering, decreases in the MF spectral profile may indicate periods requiring mental effort during HCI; however, these can be caused not only by usability problems of the software (bugs). Mental effort can also be attributed to imperfect training of the user (user error), or it can be caused by normal accompanying effort (e.g., mental effort is required by learning tasks when using e-learning software or doing creative tasks), or sometimes it can also be caused by other mental processes, independent from software use. The differentiation between them can be supported with statistics [36]. The interpretation must be based on exploring the keyboard and mouse event logs and the captured screens. Interviews can also help. However, understanding the real mechanisms underlying the interactions remains difficult.

The results in this paper show new possibilities that can be used in future assessments of HCI. The diagnostic value of these new possibilities should be studied. Naturally, further validation and exploration of the boundaries of the possibilities are also necessary.

Involving more channels is a possible way to improve this methodology. If different channels indicate the same attributes (e.g., if the HPV profile curve has a decrease, the video images of the posture, gestures, and facial expression show mental effort, and the person confirms this during the interview), their synergy can improve the interpretations.

In other cases, the different channels can complete each other. Measuring SC is a new opportunity to modulate the results. This new opportunity can initialize new studies.

REFERENCES

1. International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC). Software engineering—product quality (Standard No. ISO/IEC 9126:1991). Geneva, Switzerland: ISO; 1991.
2. 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.
3. Mulder G, Mulder-Hajonides van der Meulen WREH. Mental load and the measurement of heart rate variability. Ergonomics. 1973;16:69–83.
4. Pagani M, Lombardi F, Guzzetti S, Rimoldi O, Furlan R, Pizzinelli P, et al. Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympathovagal interaction in man and conscious dog. Circ Res. 1986;59:178–93.
5. Láng E, Szilágyi N. Significance and assessment of autonomic indices in cardiovascular reactions. Acta Physiol Hung. 1991;78(3):241–60.
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. Láng E, Szilágyi N, Métneki J, Czeizel E, Ádám G. A psychophysiological method for discrimination of MZ and DZ twins: a study based on impedance cardiography. In: Proceedings of the 6th International Congress on Twin Studies, Rome, Italy. 1989.

13. Láng E, Caminal P, Horváth G, Jané R, Vallverdu M, Slezsák I, et al. Spectral analysis of heart period variance (HPV)—a tool to stratify risk following myocardial infarction. J Med Eng Technol. 1998;22(6):248–56.

14. Silbernagl S, Despopoulos A. Color atlas of physiology. 6th ed. Stuttgart, Germany: Thieme; 2009.

15. Akaike H. Fitting autoregressive models for prediction. Ann Inst Statist Math. 1969;21:243–7.

16. Kay SM, Marple SL. Spectrum analysis—a modern perspective. Proceedings of the IEEE. 1981;69(11):1380–419.

17. Gray AH, Wong DY. The Burg algorithm for LPC speech analysis synthesis. IEEE Trans Acoust Speech Signal Process. 1980;28(6):609–15.

18. Andreassi JL. Psychophysiology human behavior and psychological response. 4th ed. Mahwah, NJ, USA: Erlbaum; 2000.

19. Luu Ph, Frank R, Kerick S, Tucker DM. Directed components analysis: an analytic method for the removal of biophysical artifacts from EEG data. In: Schmorrow DD, Estabrooke IV, Grootjen M, editors. Augmented cognition, HCII 2009 (LNAI 5638). Berlin, Germany: Springer; 2009. p. 411–6.

20. Brouwer A-M, Hogervorst MA, Herman P, Kooi F. Are you really looking? Finding the answer through fixation patterns and EEG. In: Schmorrow DD, Estabrooke IV, Grootjen M, editors. Augmented cognition, HCII 2009 (LNAI 5638). Berlin, Germany: Springer; 2009. p. 329–38.

21. Lee H, Lee J, Seo S. Brain response to good and bad design. In: Jacko JA, editor. Human–Computer Interaction, Part 1, HCII 2009 (LNCS 5610). Berlin, Germany: Springer; 2009. p. 111–20.

22. Huang R-S, Jung T-P, Makeing S. Tonic changes in EEG power spectra during simulated driving. In: In: Schmorrow DD, Estabrooke IV, Grootjen M, editors. Augmented cognition, HCII 2009 (LNAI 5638). Berlin, Germany: Springer; Proceedings; p. 394–403.

23. Srinivasan R, Thorpe S, Deng S, Lappas T, D'Zamura M. Decoding attentional orientation from EEG spectra. In: Jacko JA, editor. Human–Computer Interaction, Part 1, HCII 2009 (LNCS 5610). Berlin, Germany: Springer; 2009. p. 176–83.

24. Mandryk RL, Atkins MS, Inkpen KM. A continuous and objective evaluation of emotional experience with interactive play environments. In: Proceedings of CHI2006—Conference on Human Factors in Computing Systems. New York, NY, USA: ACM Press; 2006. p. 1027–36.

25. Park B. Psychophysiology as a tool for HCI research: promises and pitfalls. In: Jacko JA, editor. Human–Computer Interaction, Part 1, HCII 2009 (LNCS 5610). Berlin, Germany: Springer; 2009. p. 141–8.

26. Obinata G, Tokuda S, Fukuda K, Hamada H. Quantitative evaluation of mental workload by using model of involuntary eye movement. In: Harris D, editor. Engineering Psychology and Cognitive Ergonomics, HCII 2009 (LNAI 5639). Berlin, Germany: Springer; 2009. p. 223–32.

27. Oliveira FTP, Aula A, Russel DM. Discriminating the relevance of web search results with measures of pupil size. In: Proceedings of CHI2009—Conference on Human Factors in Computing Systems. New York, NY, USA: ACM Press. 2009. p. 2209–12.
28. Tullis T, Albert B. Measuring the user experience. Collecting, analyzing, and presenting usability metrics. Amsterdam, The Netherlands: Elsevier; 2008.
29. Komlodi A, Hercegfi K. Cultural differences in information behavior. In: Proceedings of HCIR 2009: 3rd Workshop on Human-Computer Interaction and Information Retrieval. New York, NY, USA: ACM Press; 2009; p. 78–81.
30. Komlodi A, Hercegfi K. Exploring cultural differences in information behavior applying psychophysiological methods. In: Proceedings of CHI2010—Conference on Human Factors in Computing Systems. New York, NY, USA: ACM Press; 2010. p. 4153–8 .
31. Izsó L, Hercegfi K. HCI Group of the Department of Ergonomics and Psychology at the Budapest University of Technology and Economics. In: Extended Abstracts of CHI2004—Conference on Human Factors in Computing Systems. New York, NY, USA: ACM Press; 2004. p. 1077–8.
32. Hercegfi K, Kiss OE, Bali K, Izsó L. INTERFACE: assessment of human–computer interaction by monitoring physiological and other data with a time-resolution of only a few seconds. In: Ljungberg J, Andersson M, editors. ECIS 2006: 14th European Conference on Information Systems. 2006. p. 2288–99.
33. Hercegfi K, Kiss OE. Assessment of e-learning material with the INTERFACE system. In: Bernath U, Szűcs A, Tait A, Vidal M, editors. Distance and e-learning in transition. Hoboken, NJ, USA: Wiley; 2009. p. 645–57.
34. Hercegfi K, Pászti M, Tóvölgyi S, Izsó L. Usability evaluation by monitoring physiological and other data simultaneously with a time-resolution of only a few seconds. In: Jacko JA, editor. Human–Computer Interaction, Part 1, HCII 2009 (LNCS 5610). Berlin, Germany: Springer; 2009. p. 59–68.
35. Laufer L, Németh B. Predicting user action from skin conductance. In: Proceedings of IUI 2008—13th International Conference on Intelligent User Interfaces. New York, NY, USA: ACM Press; 2008. p. 357–60.
36. Izsó L. Discrimination between design errors and user errors by binomial test. Behav Inf Technol. 2000;19(5):379–84.