Heart Rate Variability Monitoring during Human-Computer Interaction

Károly Hercegfi

Department of Ergonomics and Psychology,
Budapest University of Technology and Economics
Magyar tudósok krt. 2, Bldg. Q
H-1117 Budapest, Hungary
hercegfi@erg.bme.hu

Abstract: This paper outlines a Heart Rate Variability (HRV)-based method applicable to Human-Computer Interaction (HCI) researches. After a brief overview of various psychophysiology-based empirical techniques (mainly focusing on the mental effort approach), a recent research is shown. The paper presents new results of a short, basic series of experiments, attempting to explore the boundaries of the temporal resolution of the method. The applied INTERFACE methodology is based on the simultaneous assessment of HRV and other data. The results raise hope that this methodology is potentially capable of exploring mechanisms underlying practical usability issues and identifying quality attributes of software elements – over the previously developed HRV-based methods – with a temporal resolution of only a few seconds.

Keywords: Human-Computer Interaction (HCI); software usability testing and evaluation; empirical methods; Heart Rate Variability (HRV)

1 Introduction

To assess the quality of software products, one of the most important keywords is usability. A highlighted usability factor is the mental effort required by current Human-Computer Interaction (HCI). In some definitions of usability as a quality dimension of software products – as it is laid down, e.g., in the original version of the international standard of software product evaluation (ISO/IEC 9126:1991 [1]) – the required mental effort appears not only as a metric, but it is the core of the definition: better quality means that less effort is needed for the usage of the particular software.

Evidently, in addition to subjective methods such as questionnaires, objective methods are also needed to measure mental effort (self-imposed mental work stress).

This paper focuses on the Heart Rate Variability (HRV) power spectrum as a technique for measuring the current mental effort as a function of time. In addition to this, a brief overview of other physiological channels applied to studying HCI is also presented.

This paper presents new results of a foundational research, aiming at exploring the boundaries of the HRV-based method, supporting base data for future applied research, and focusing on future usability evaluations of software.

1.1 Assessing Mental Effort via Analysing Users' HRV Power Spectra

Sometimes the Heart Rate (HR) itself is used in usability evaluations; however, it is not a sensitive measure of mental effort and thus of 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 Heart Rate Variability (HRV) is needed. Although in the literature the term "Heart Rate Variability" (HRV) is more frequently mentioned, we prefer the similar expression "Heart Period Variability" (HPV), where the periods of time between consecutive heart beats are simply the reciprocal values of the heart rates: in practice, the periods of heart beats can be analysed more directly, and they can be more expressive.

HRV is applied in various areas. Naturally, there are a series of realizations in medicine (e.g. [14] [17] [18] [19]). Software usability methods have been influenced by these techniques.

The time periods between heart beats are called RR intervals, because they are determined by the highest peaks (the so-called R peaks) of electrocardiogram (ECG) curve. (Additionally, in some papers the RR intervals are referred as the easy-to-remember "Rhythm-to-Rhythm" intervals [29].)

After analysing the variability of the RR intervals, a number of studies [11] [12] [19] [22] [25] [26] [29] [30] [32] [35] have shown that an increase in mental load causes a decrease in the so-called *mid-frequency (MF, 0.07-0.15 Hz) power band* of the Heart Period Variability (HPV) power spectrum. Focusing on this frequency band filters other peaks of the power spectrum: the typical peak in the 0.15-0.45 Hz band corresponds to the respiratory rate (called respiratory sinus arrhythmia); the peak in the 0.04-0.07 Hz band is in connection with the thermoregulatory fluctuations of the blood vessels [12] [19]. Heart rate fluctuations in the MF (0.07-0.15 Hz) power band may also reflect postural changes (via the blood pressure control of the so called baroreflex). To separate the effect of the mental load from the effect of postural changes, a ratio of the MF component around 0.1 Hz and the higher frequency respiratory component can be applied [30]. However, it is emphasised that if the participants work continuously in a sitting posture (e.g., during computer usage), and their larger muscle movements (e.g., stretching, laughing, sneezing, talking, etc.) eventually are filtered from the records (e.g., via video analysis), the MF (0.07-0.15 Hz) power band itself can characterize the mental effort sensitively enough, as is shown by the following results presented in this paper.

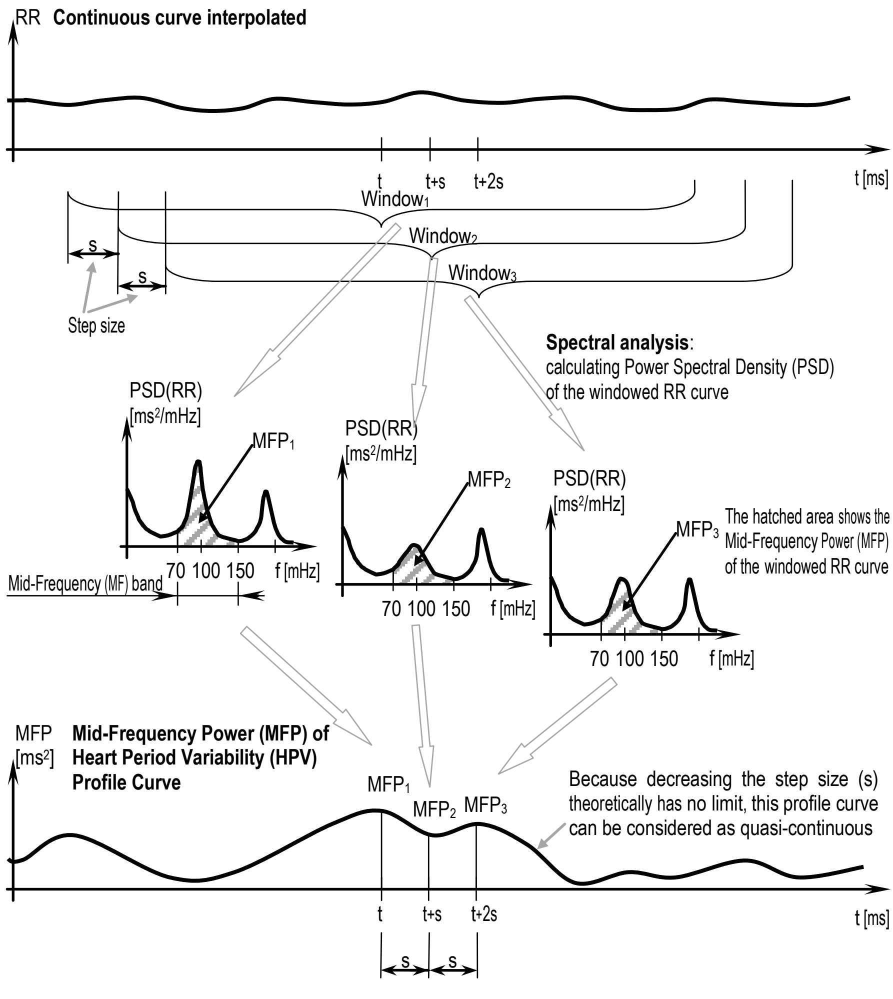

Figure 1

Calculation of the Mind-Frequency (MF) power of the Heart Rate Variability (HPV) profile curve

To assess the spectral components of the HPV power spectra, an integrated system called ISAX (Integrated System for Ambulatory Cardio-respiratory data acquisition and Spectral analysis) was developed and successfully used by Láng and her team [11] [17] [18] [19]. This equipment and the related method have been integrated into our INTERFACE system.

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 analyse only one segment at the very beginning of the time series of the RR intervals. This can be performed by applying *windowing* functions: in this way, a selected segment (frame) can be characterized by the calculated MF power of the HPV. When it is done, the frame is shifted a bit further and the spectral analysis is repeated, and so on many times, until the frame reaches the end of the time series (see Figure 1). This kind of analysis technique is embedded in the ISAX system: this windowing technique is applied by scrolling a constant-size frame by small steps (in this series of experiments, 32-second frames were windowed, and the frame was repeatedly shifted by 1-second 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 so-called *MF spectral profile curve*. As the maximal delay of detection is equal to the step size, the change in HPV in principle can be reflected in the spectral profile curve within one second.

The main advantage of our method over the previously developed HPV-based methods [25], in our opinion, is that the MF component of HPV is able to indicate changes in mental effort *within a range of several seconds* (as opposed to the earlier methods with a resolution of tens of seconds at best). This feature was achieved by the following:

- Applying an appropriate windowing data processing technique using the Hamming windowing function to decrease the aperture effect and improve the spectrum image.
- Applying an all-pole auto-regressive model (instead of, e.g., spectrum analysis based on the simple Fourier algorithm) with built-in recursive Akaike's Final Prediction Error criteria [2] [13] and a modified Burg’s algorithm [5]. The autoregressive model can already be used in only one cycle, and can give well-established results in the case of 2-3 cycles, contrary to the Fourier algorithm, which requires 8-10 cycles. This means that, if we focus on 0.1 Hz, a 10-second frame can give some results, and a 20- or 30-second frame ensures good result. If the low boundary of the MF band (0.07 Hz) is taken into consideration, a 15-second frame can give some results, and a 30- or 45-second frame ensures good result.
- Creating the above-mentioned spectral profile curve, based on overlapping windows, by finding the best compromise between the spectral and the temporal resolution. A wider window allows for better spectral quality; however, it blurs the effects of longer period.

1.2 Other Physiological Channels Applied to Studying HCI

There exist several other physiology-based techniques to analyse HCI.

A part of these aims to measure actual mental effort, while others aim to identify emotional aspects of HCI. Emotions can represent an independent dimension from the mental effort; however, its importance can also be similarly high in the HCI practice.

Changes in the electrical characteristics of the skin (the so-called Electrodermal Activity – EDA) can be evoked by various physical and emotional stimuli. In our practice, the parameters derived from *Skin Conductance (SC)* responses, especially the Alternating Current (AC) component of the SC, are used.

Although there are other techniques for measuring mental effort and emotions, either they are more difficult to evaluate and more disturbing for the participants (e.g. the Electroencephalograph – EEG), or they give an overall, averaged indicator for a relatively long period of time, from minutes to hours (for example, the visual critical flicker frequency (CFF) and the practical applications of certain bio-chemical measures).

EEG requires sophisticated set of electrodes and the participants experience it as more disturbing. Furthermore, it results in much more complex curves. Various effects have to be filtered from the data, such as the effects of the eye blinks, among many others [3] [23] [35]. Naturally, EEG can explore many more aspects of mental effort than the ECG can; however, if only a single metric of the mental effort is targeted, ECG is a simpler and more preferred method.

Applying EEG can be a potential direction of *further developments* of our methodology: not to simply identify mental effort, but (1) to identify more complex mental or emotional state patterns (using complex methods to analyse the complex curves [4] [21]), or (2) to attempt to localize the active brain regions (using more than 20 [35] or 128- or 256-channel Dense Array EEG (dEEG) [9] [33]).

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 provides a very good indicator of tension in an individual due to jaw clenching. On the face, EMG has been used to distinguish between positive and negative emotions. EMG activity over the brow (frown muscle) region is lower and EMG activity over the cheek (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 about 5 mm) and the closeness of the muscles of the different mimic functions, the electrodes have to be positioned extremely carefully [31]. Furthermore, the participants experience the electrodes on the face or head again as more disturbing than the electrodes on the fingers of the non-dominant hand measuring Skin Conductance (SC). Thus, SC is applied as the simpler and preferred method to identify emotional reactions, instead of the EMG's potential capability of differentiating positive and negative emotions.

Measuring mental effort by visual *Critical Flicker Frequency (CFF)*, and in a biochemical way (measuring, e.g., the *cortisol level of the saliva*) have also been applied by members of our team [12]. However, these methods give only an overall, "washed-together", averaged indicator for a relatively long period of time, from several minutes to hours – this is not the fine temporal resolution targeted by our INTERFACE methodology.

Eye-tracking is a promising direction of *further developments* of our methodology: (1) 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 (2) it can be analysed deeper, deriving parameters referring to the state of the nervous system [27].

Pupillometry (measuring the current diameter of the pupil) is a measurement option that is often accomplished with eye-tracker equipment. It is reflective of both the mental effort and the emotions of the user [3] [28] [34]. It can be capable of validating the other physiological channels of our methodology.

Eye-tracking and pupillometry are used in our *ongoing INTERFACE research* [15] [16].

2 Applied Methods

2.1 The INTERFACE Methodology

A complex methodology was developed at the Budapest University of Technology and Economics, by Izsó and his team [6] [7] [8] [10] [11] [12].

Figure 2 shows the conceptual arrangement of the INTERFACE (INTegrated Evaluation and Research Facilities for Assessing Computer-users' Efficiency) workstation.

The advantage of the methodology lies in its capability of recording continuous on-line data characterizing the user's current mental effort derived from *Heart Period Variability (HPV)* simultaneously and synchronized with other characteristics of Human-Computer Interaction (HCI), such as screen captures and a log of all mouse and keyboard use input. In this way, a detailed picture can be obtained which can serve, after a series of careful considerations, as a basis for a deeper understanding and interpretation of the psychological mechanisms underlying HCI.

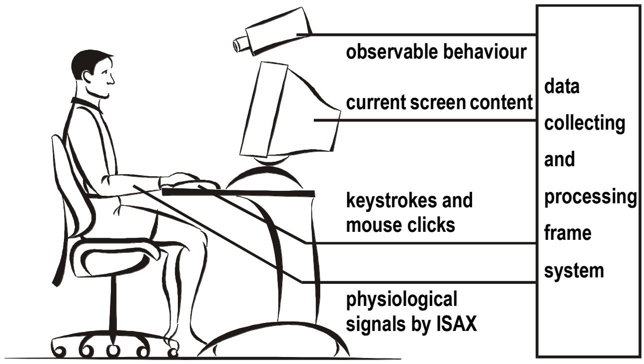

Figure 2

Conceptual arrangement of the INTERFACE software testing workstation

The INTERFACE simultaneously investigates the following:

- Users' observable actions and behaviour
  - keystroke and mouse events;
  - video record of the current screen content;
  - video records of users' behaviour: (1) facial expression, (2) posture and gestures.
- Psychophysiological parameters
  - power spectrum of Heart Period Variability (HPV), regarded as an objective measure of current mental effort – we have applied this measure successfully for more than 15 years [6] [7] [8] [11] [12];
  - in some cases, this system is completed by other physiological channels, such as Skin Conductance (SC) [8], and/or pupillometry and eye-tracking [15] [16] as well.

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

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

2.2 Experimental Arrangement and Participants

The experiments presented here were carried out at the Budapest University of Technology and Economics.

Three ECG electrodes were placed on the user's torso:

- the exploring (positive) electrode on the 7^{th} or 8^{th} rib (below the left nipple);
- the indifferent (negative) electrode high up on the right side of sternum (breastbone), i.e. on the right side of the manubrium of the sternum, close to the right clavicle (collar-bone), or in the left side of the right infraclavicular fossa;
- the ground electrode on the 8^{th} or 9^{th} rib on the left median auxiliary line.

(Depending on the body form and structure of the participant, other ECG electrode locations can also be selected to maximize the magnitude of the R wave and minimize the artefacts caused by movements).

The signals were recorded by the afore-mentioned ISAX system.

Two notebooks were used: one for the participant, one for the experimenter. Two video cameras (USB web cameras) were applied (one of them with face-tracking capabilities). Instances of the Virtual Dub 1.9 software were applied as video capture software. The video capture of the screen content was realized by the Hypercam 2.1 screen recorder software. The applied keyboard and mouse event logger software, the experimenter's notation software, and the software and hardware elements of the frame system of the synchronization were developed by our team at the Budapest University of Technology and Economics.

For studying the HCI, two pieces of software were used by the participants:

- An arcade game: the 8^{th} episode of the popular YetiSports series (www.yetisports.org), called "Jungle Swing". It had already been applied by colleagues at our department to other experiments [20].
- A game to exercise arithmetic: the Raindrops game of the Lumosity "brain training" web site (www.lumosity.com).

Both games run in web browser – Microsoft Internet Explorer 8 was used. (Because of these circumstances, in this series of experiments, not all the mouse and keyboard events were recorded by our logger software; however, based on the captured screen video, it was possible to reconstruct the mouse and keyboard events).

After three pilot sessions, ten regular sessions were recorded. Seven of the participants of the regular sessions were female, three of them were male. Their ages were 20 to 35.

2.3 The Viewer Screen of the INTERFACE Software

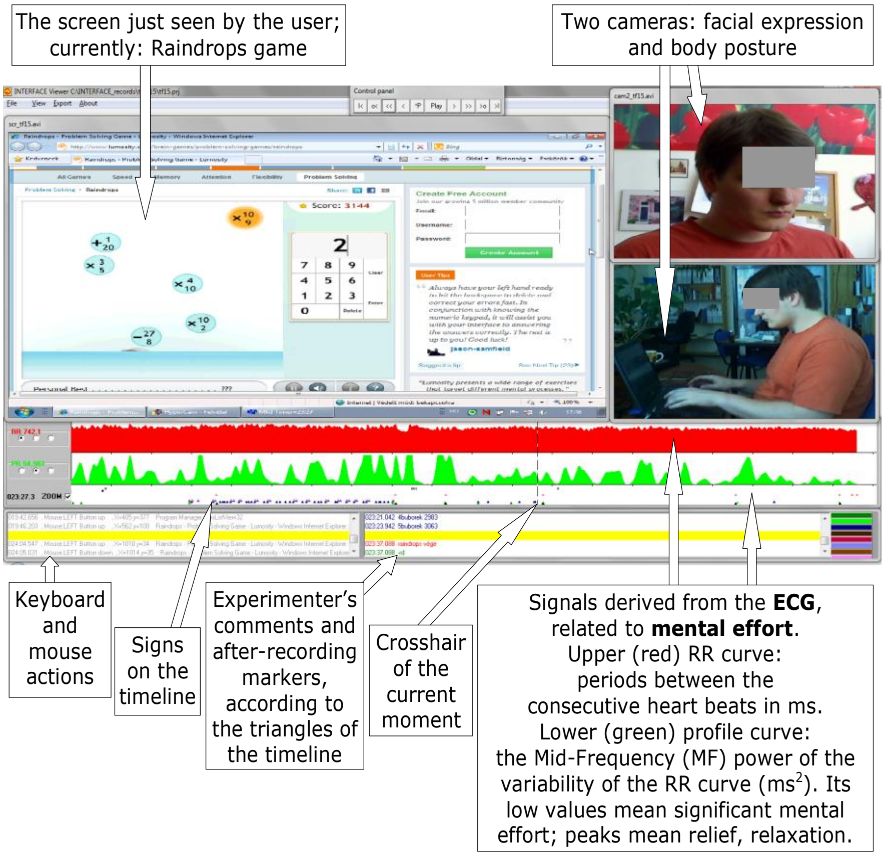

Figure 3

The INTERFACE Viewer screen with a record of the series of experiments. As it can be seen, the user at the selected point is exerting significant mental effort – it is shown by the facial expression and posture of the user and the low value of the last profile curve of the Mid-Frequency (MF) power of the Heart Period Variability (HPV) at the cross-hair. (The curves displayed in the current window show the history of 24 minutes.)

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 current series of experiments. This figure also shows the typical pattern of mental effort observable both on the HPV curve and in the video images.

2.4 Schedule of Each Session

Each session had the following schedule:

I At the beginning of each session, we performed a “calibration” phase

1. First, the participant was asked to *relax* for approximately two minutes. The instructions of the relaxation periods were always the same:
   - seating themselves in a comfortable posture, without any movement,
   - keeping their eyes open,
   - trying to be thinking of nothing, in spite of our knowing it is not trivial for people untrained to this (at least trying to avoid to think specific items),
   - calming them down, ensuring them that there are not good and bad personal results: we have no expectations, we would like to experience only some differences between this period and the next one.
2. The relaxation was followed by a two-minute artificially induced high mental effort exercise: mental arithmetic. However, after the instructions, an *anticipation* period (20 seconds) followed: forcing the participant to wait for the next, mental effort demanding task. Both the anticipation and the mental arithmetic task were controlled by a PowerPoint slide show. The preliminary instructions of the *mental arithmetic* periods were also always consequent:
   - no movements,
   - no speech, no aloud counting, and neither voiceless movement of mouth,
   - after giving the participants a starting number by the slide show (in these cases: 11558), the participants immediately had to count backwards by 7s (it is really difficult and requires a high level of mental effort);
   - two minutes later, the result of the counting was asked by the slide show – however, the actual result of the counting was not really important in itself; the only goal was to artificially generate mental effort.

     It was an important element of the experiment design that the mental arithmetic periods were immediately followed by *relief* periods: after answering the result of the mental effort demanding task, there came 10-20-second periods without anything to do (without new instructions, but only with some confirming smile).

II Playing with the mentioned *YetiSports computer game*

This game was selected because of the simplicity of the interaction: the user has to use only one mouse button and just click at the right time. In this 8th episode called Jungle Swing, the yeti hangs and swings around branches of trees. By pressing the mouse button, the all-the-time-swinging yeti swings away from the branch tangentially. The goal of the game is to help the yeti to get to the highest branches of a tree. To do so, one must figure out the best timing for the yeti; when to leave the branch to jump to the next branch where it starts to swing around the new branch, and from where it can jump further and further again [20].

If the user (player) fails to do so, the yeti will fall down, until it is able to catch a branch or until it falls into the water underneath. After falling down into the water, a message box appears; then when the player clicks the button, the start screens of the game are displayed. It means that each fall into the water is followed by the same five clicks to restart the game, and to make a new attempt. This alternation of concentration-requiring attempts and drilled, easy clickstreams to restart the game has a role in this series of experiences.

The game looks easy; however it is very hard for anyone who is a novice. The jumps are difficult to make; they need very precise timing. To avoid severe frustration that could influence the subsequent tests, we ended the game at the first instance of the participant making at least four successful jumps in a row.

III Playing with the mentioned *Raindrops arithmetic exercise game*

This game is an arithmetic test, where participants have to solve short equations, such as 4+9, or 16/4, within a given time period. Each equation appears in a raindrop, falling towards the bottom of the screen, and has to be solved before it reaches the ground or the water level rises. If the water level reaches the top of the screen, the game ends. The equations become more and more difficult (additions and subtractions at first, then multiplications and divisions later), and over time they become more frequent.

2.5 HRV Analysis, Statistical Analysis

Data recording of the ECG peaks was performed by the ISAX equipment, as was mentioned in the introduction. Data processing of the collected raw ECG peaks and the power spectrum analysis were performed by the ISAX software. Creating the mentioned spectral profile curve, the following parameters were applied:

- frequency band: 0.07-0.15 Hz (MF);
- size of the windowing frame: 32 sec;
- steps (shifts) of the windowing frame: 1 sec.

Because of the low number of participants, the normality of the distribution of measured parameters cannot be proved. Therefore a non-parametric statistical method, the Wilcoxon Signed Ranks Test, was used to test the differences.

Statistical analyses were performed using the IBM SPSS Statistics, version 19.0.

3 Results

3.1 Differences between the Mental Effort Values Measured during the Different Tasks

The curves shown in the Figure 4 were recorded during the 6th session.

In Figure 4 the upper (red) curve represents the RR values (heart periods), and the bottom (green) one displays the Mid-Frequency (MF) power profile curve of Heart Period Variability (HPV).

During relaxation, the MF component of the HPV increases, so the profile curve runs relatively high (and, naturally, the RR curve has zigzags). In the case of "perfect" relaxation, the profile curve should be continuously high. However, this is not expected in this experimental situation: the participants were not trained to use advanced relaxation techniques.

Then the anticipation section follows.

During the mental arithmetic exercise, the RR curve definitely gets smoother, and consequently the profile curve also gets significantly lower. In this figure, the profile curve can be considered low, especially in comparison with the other sections.

After the "calibration" tasks, the participants are relieved.

In Figure 4, the curves show that this participant really could relax. (The profile curve in Figure 4 has its highest peak in the middle of the section of relaxation.) However, in this series of experiments, most of the participants could not relax well: they were just wired, and they felt the relaxation as a serious task, and it even caused a certain task load. However, the mental arithmetic exercise usually results in the excepted low curve, which can be used as a baseline. Furthermore, in most cases, during the mentioned short period of relief, the participants get more relaxed than during the conscious, intended relaxation: the MF of HPV profile curves have their highest peaks here (this is the so called "rebound" phenomenon). And, in most cases, the anticipation period meant more relaxed phases than the original relaxation task.

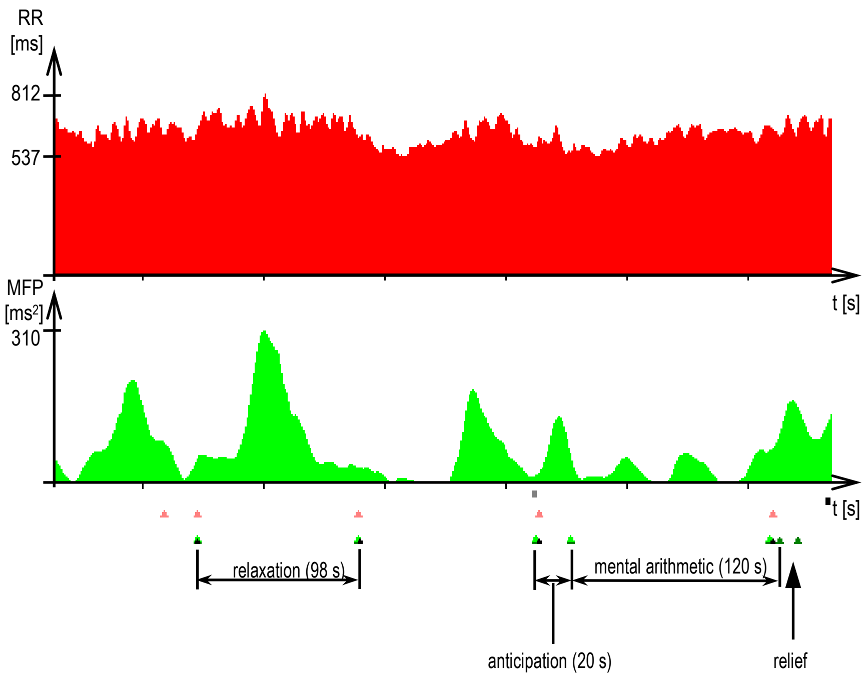

Figure 4

The typical pattern of the relaxation and mental arithmetic periods in the case of the 6th participant. The expressive visualisation style of the INTERFACE Viewer software is applied. The upper (red) curve represents the RR intervals (heart periods), and the bottom (green) one displays the Mid-Frequency (MF) power profile curve of the Heart Period Variability (HPV).

Figure 5 compares the values of the MF power of the HPV profile curves during the above mentioned "calibration" phase and the playing phases.

Studying the results of the ten sessions, the MF power of the HPV profile curve values looks higher during the relaxation periods than during mental arithmetic. However, the Wilcoxon Signed Ranks Test has not proven the difference (sig. 0.254>>0.05). It may be caused by the mentioned effects (or, naturally, it was influenced by other effects). Probably a sample of a higher number of participants could give significant result also with these circumstances.

However, the differences between the mental arithmetic task and the anticipation period, and the mental arithmetic task and the relief are significant: the Wilcoxon test results *sig. 0.019* and ***sig. 0.005***.

In Figure 5 the YetiSports looks easier than the Raindrops, and the Raindrops looks easier than the pure mental arithmetic test of the "calibration" phase. However, these differences are not significant in term of statistics. In comparison of the two games, the Wilcoxon test results sig. 0.069>0.05.

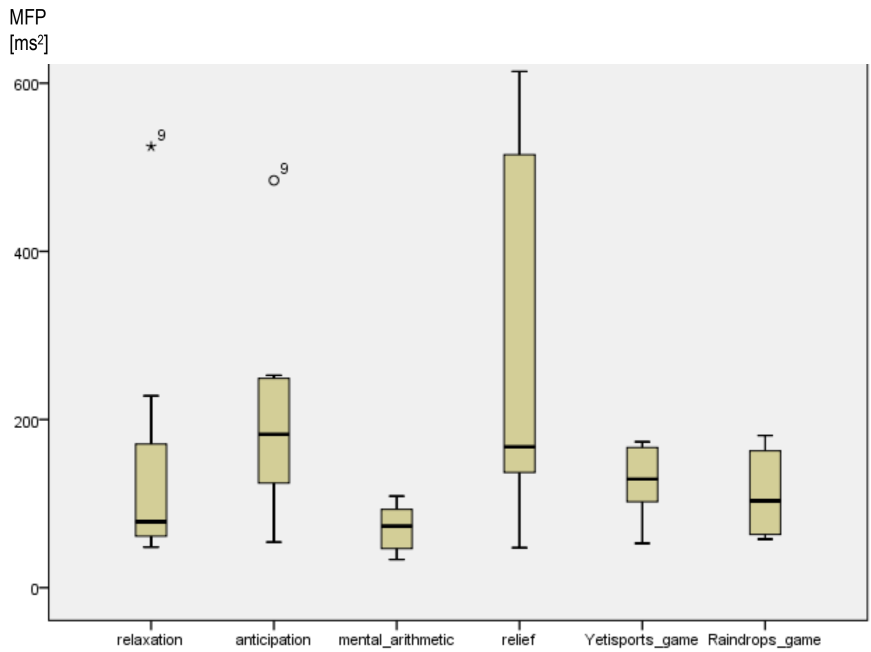

Figure 5

Boxplots[1] of the Mid-Frequency (MF) power of Heart Period Variability (HPV) values during the different tasks and periods

3.2 A Highlighted Result of the Current Series of Experiments: Proofing the Differences between Short Periods

As was mentioned earlier, the YetiSports game is very difficult. During the 5-15 minutes of play with this game, each player had 11-35 attempts. It means that most of the attempts were really short; a number of them contained only a single jump (of course, an unsuccessful jump, and a fall). Other, rare attempts contained 5-10 jumps. The average length of an attempt is 16.4 sec (min. 4.5, max. 79.4, std. dev.: 11.0).

[1] Boxplots are applied as usual. The dark lines in the middle of boxes are the medians. The bottoms of the boxes indicate the first quartiles, the tops of the boxes represent the 3rd quartiles. The T-bars (the inner fences or whiskers) extend to 1.5 times the height of the box, or, if no case/row has a value in that range, to the minimum or maximum values. The circles and asterisks are outliers – these are defined as values that do not fall in the inner fences. Asterisks are extreme outliers – these represent cases/rows that have values more than three times the height of the boxes.

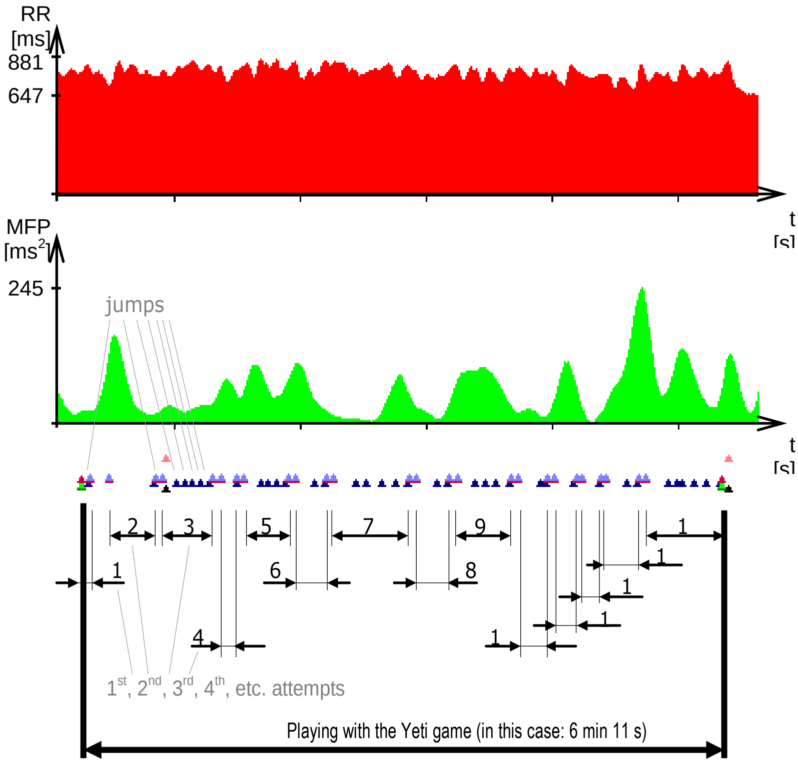

Figure 6

The typical pattern of the playing with the YetiSports arcade game in the case of the 10th participant. The upper (red) curve represents again the RR values (heart periods), and the bottom (green) one displays the Mid-Frequency (MF) power profile curve of the Heart Period Variability (HPV). In this case, during the play, the user had 14 attempts. Some of the attempts meant only one jump, and the yeti immediately fell down (during the 1st, 2nd, 4th, 11th attempts). Some of the attempts were more successful (e.g. the 3rd attempt contained 5 jumps, the 14th attempt contained 6 jumps). The periods between two attempts are short (in this case the average was 4.7 sec, the shortest one was 3.2 sec). As it can be seen, these periods (simply clicking 5 times after each fall down to restart the game) usually do not show mental effort: most of them are followed by peaks of the profile curve.

As was mentioned earlier, after each fall down, 5 clicks are required to restart the game. These are simple, effortless clicks, each time using the same buttons. These periods between two consecutive attempts are short. The average length of a "break-time" between two consecutive attempts is ***6.2 sec*** (min. 3.2, max. 13.9, std. dev.: 2.4).

The typical pattern of alternation of these two types of periods are shown in Figure 6.

The alternation of the MF power of HPV is significant, ***in spite of the very short periods***!

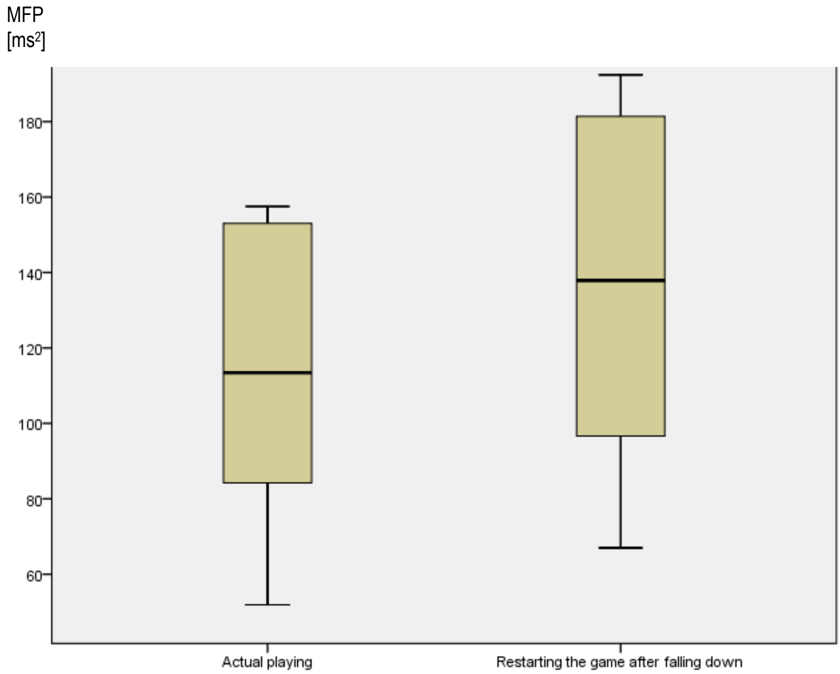

Figure 7
Comparing the mental effort required by the attempts of playing with the YetiSports arcade game (actual playing) and the easier "breaking-time" periods between the consecutive attempts (simply clicking 5 times after each fall to restart the game). (These periods are exactly defined in the main text.) The boxplots[1] show the significant difference (*sig. 0.006*) between the aggregated Mid-Frequency (MF) power of Heart Period Variability (HPV) values during these periods.

The comparison of the aggregated values of the "actual playing" periods and the "breaking-time" periods is shown in Figure 7. The definitions of these periods were the following:

- Actual playing period = from last click on the button right before the appearance of the yeti until the disappearance of the falling yeti.
- Breaking-time period = from the disappearance of the falling yeti until the last click on the button right before the next appearance of the yeti.

The Wilcoxon Signed Ranks Test has proven the difference: ***sig. 0.006*** << 0.05.

Discussion

Based on the results presented here as well as in other related papers, it can be stated that the INTERFACE methodology already in its present form is capable of identifying the relatively weak points of the HCI. With this methodology, it was possible to study events occurring during the HCI *in such a high temporal resolution* and with such objectivity that would not have been possible using other methods presently known to us. (An overview of other methods was presented in the Introduction.)

The applied Heart Period Variability (HPV) profile function integrated into the INTERFACE system seems to be a potentially powerful tool for monitoring events in very narrow time frames. The theoretical establishment of this feature was explained in section 1.1, and the results of this series of experiments have empirically proved the capability of identifying differences of 6.2 sec periods).

Naturally, analysis of the synchronized records of physiological data and videos together is a must. By the help of the video records of the cameras, for example, the artefacts of HPV caused by large muscle movements (such as stretching) can be filtered out; these peaks of the profile curve cannot be interpreted mechanically as decrease of mental effort. In other cases, the peaks of the profile curve can indicate relaxed periods during easy software-usage, or they can indicate “giving up” (“no coping any more”) situations; they can be differentiated on the basis of the efficiency of the activity of the participant. (Even if there are objective performance metrics, however, this interpretation must be carried out very carefully.)

After appropriate artefact filtering, decreases in the MF spectral profile may indicate periods requiring mental effort during the HCI – however, these can be caused not only by usability problems of the software (software flaw). Mental effort can also be attributed to the imperfection of the training of the user (user error), or it can be caused by normal accompanying effort (e.g., naturally, mental effort is required in learning tasks when using e-learning software or accomplishing creative tasks), or sometimes it can also be caused by other thinking process of the participant, independent from the actual software use. The interpretation must be based on exploring the recorded keyboard and mouse event logs and the captured screens. It can also be helped by interviews. However, understanding the real mechanisms underlying the interactions still remains difficult.

The results presented in this paper show new possibilities that can be applied in future assessments of HCI. The diagnostic value of these new possibilities may be explored during future studies. Naturally, further validation and exploration of the boundaries of the possibilities are also needed.

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 shows a decrease, the video images of the posture, gestures, and facial expression show mental effort, and the person himself confirms it during the interview), the synergy between them can help in forming interpretations.

In other cases, the different channels can complete each other. For example, measuring the Skin Conductance (SC) and/or pupil size are a new opportunities to modulate the results. These new opportunities can initialize new studies.

Acknowledgement

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

References

[1] International standard: ISO/IEC 9126:1991. Software engineering – Product quality; 1991

[2] Akaike, H.: Fitting Autoregressive Models for Prediction. Ann. Inst. Statist. Math. 1969;21:243-7

[3] Andreassi, J. L.: Psychophysiology Human Behavior and Psychological Response. 4th ed. Mahwah: Lawrence Erlbaum Associates; 2000

[4] Brouwer, A-M., Hogervorst, M. A., Herman, P., Kooi, F.: Are You Really Looking? Finding the Answer through Fixation Patterns and EEG. In: Shmorrow et al., editors: Augmented Cognition, HCII2009, LNAI 5638. Springer; 2009; pp. 329-338

[5] Gray, A. H., Wong, D. Y.: The Burg Algorithm for LPC Speech Analysis Synthesis. IEEE Tr. ASSP. 1980; ASSP-28(6):609-15

[6] Hercegfi, K., Kiss, O. E., 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: Proceedings of ECIS 2006 – 14th European Conference on Information Systems; 2006 June 12-14, Göteborg, Sweden. ECIS Standing Comm.; 2006; pp. 2288-99

[7] Hercegfi, K., Kiss, O. E.: 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: John Wiley & Sons; Chapter 45, pp. 645-57

[8] 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: Proceedings of HCII2009 (Human-Computer Interaction International 2009); 2009 July 14-24; San Diego, CA, USA 2009; Part I. LNCS 5610. Springer; 2009, pp. 59-68

[9] Huang, R-S., Jung, T-P., Makeing, S.: Tonic Changes in EEG Power Spectra during Simulated Driving. In: Shmorrow et al., editors: Augmented Cognition, HCII2009, LNAI 5638. Springer; 9Proceedings; pp. 394-403

[10] 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; 2004 Apr 24-29, Vienna, Austria. ACM Press; 2004; pp. 1077-8

[11] Izsó, L., Láng, E.: Heart Period Variability as Mental Effort Monitor in Human Computer Interaction. Behaviour & Information Technology. 2000;19(4):297-306

[12] Izsó, L.: Developing Evaluation Methodologies for Human-Computer Interaction. Delft: Delft University Press; 2001

[13] Kay, S. M., Marple, S. L.: Spectrum Analysis - A Modern Perspective. Proceedings of the IEEE. 1981;69(11):1380-419

[14] Khoór, S., Kecskés, I., Kovács, I., Verner, D., Remete, A., Jankovich, P., Bartus, R., Stanko, N., Schramm, N., Domijan, M., Domijan, E.: Heart Rate Analysis and Telemedicine: New Concepts & Maths. Acta Polytechnica Hungarica. 2008;5(1):135-45

[15] Komlodi, A., Hercegfi, K.: Cultural Differences in Information Behavior. In: Proceedings of HCIR2009: 3rd Workshop on Human-Computer Interaction and Information Retrieval; 2009 Oct 23, Washington, D.C., USA. ACM Press; 2009; pp. 78-81

[16] Komlodi, A., Hercegfi, K.: Exploring Cultural Differences in Information Behavior Applying Psychophysiological Methods. In: Proceedings of CHI2010 – Conference on Human Factors in Computing Systems; 2010 Apr 10-15; Atlanta, GA, USA. ACM Press; 2010; pp. 4153-8

[17] Láng, E., Caminal, P., Horváth, G., Jané, R., Vallverdu, M., Slezsák, I., Bayés de Luna, A.: Spectral Analysis of Heart Period variance (HPV) – a Tool to Stratify Risk Following Myocardial Infarction. Journal of Medical Engineering and Technology. 1998;22(6):248-56

[18] 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. Proceedings of the 6th International Congress on Twin Studies; 1989, Rome, Italy

[19] Láng, E., Szilágyi, N.: Significance and Assessment of Autonomic Indices in Cardiovascular Reactions. Acta Physiologica Hungarica 1991;78(3):241-60

[20] Laufer, L., Németh, B.: Predicting User Action from Skin Conductance. In: Proceedings of IUI 2008 – 13th International Conference on Intelligent User Interfaces; 2008 Jan 13-16; Gran Canaria, Spain. ACM Press; 2008; pp. 357-60

[21] Lee, H., Lee, J., Seo, S.: Brain Response to Good and Bad Design. In: Proceedings of HCII2009 (Human-Computer Interaction International 2009) 2009 July 14-24; San Diego, CA, USA. Part I, LNCS 5610. Springer; 2009; pp. 111-20

[22] 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; 2006 Nov 2-4, Banff, Canada. ACM Press; 2006 pp. 364-71

[23] Luu, Ph., Frank, R., Kerick, S., Tucker, D. M.: Directed Components Analysis: An Analytic Method for the Removal of Biophysical Artifacts from EEG Data. In: Shmorrow et al., editors: Augmented Cognition, HCII2009, LNAI 5638. Springer; 2009; pp. 411-416

[24] Mandryk, R. L., Atkins, M. S., Inkpen, K. M.: A Continuous and Objective Evaluation of Emotional Experience with Interactive Play Environments. In:

Proceedings of CHI2006 - Conference on Human Factors in Computing Systems; 2006 Apr 22-26, Montreal, Canada. ACM Press; 2006; pp. 1027-36

[25] Mulder, G., Mulder, L. J. M., Meijman, T. F.,Veldman, J. B. P., Roon, A. M.: A Psychophysiological Approach to Working Conditions. In: Backs, R. W., Boucsein, W., editors: Engineering Psychophysiology: Issues and Applications. Mahwah: Lawrence Erlbaum Associates; 2000; Chapter 4, pp. 79-110

[26] Mulder, G., Mulder-Hajonides van der Meulen, W. R. E. H.: Mental Load and the Measurement of Heart Rate Variability. Ergonomics. 1973;16:69-83

[27] Obinata, G., Tokuda, S., Fukuda, K., Hamada, H.: Quantitative Evaluation of Mental Workload by Using Model of Involuntary Eye Movement. In: Harris, D. editors: Engineering Psychology and Cognitive Ergonomics, HCII2009, LNAI 5639. Springer; 2009; pp. 223-32

[28] Oliveira, F. T. P., Aula, A., Russel, D. M.: Discriminating the Relevance of Web Search Results with Measures of Pupil Size. Proceedings of CHI2009 - Conference on Human Factors in Computing Systems; 2009 Apr 4-9 Boston, MA, USA. ACM Press. 2009; pp. 2209-12

[29] Orsilia, R., Virtanen, M., Luukkaala, T., Tarvainen, M., Karjalainen, P., Viik, J., Savinainen, M., Nygard, C-H.: 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

[30] Pagani, M., Lombardi, F., Guzzetti, S., Rimoldi, O., Furlan, R., Pizzinelli, P., Sandrone, G., Malfatto, G., Dell'Orto, S., Piccaluga, E., Turiel, M., Baselli, G., Cerutti, S. and Malliani, A.: 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

[31] Park, B.: Psychophysiology as a Tool for HCI Research: Promises and Pitfalls. In: Proceedings of HCII2009 (Human-Computer Interaction International 2009). 2009 July 14-24; San Diego, CA, USA; Part I, LNCS 5610. Springer; 2009; pp. 141-48

[32] Rowe, D. W., 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; 1998 Apr 18-23, 1998; Los Angeles, CA, USA. ACM Press; 1998, pp. 18-23

[33] Srinivasan, R., Thorpe, S., Deng, S., Lappas, T., D'Zamura, M. (2009): Decoding Attentional Orientation from EEG Spectra. In: Proceedings of HCII2009 (Human-Computer Interaction International 2009) 2009 July 14-24; San Diego, CA, USA; Part I, LNCS 5610. Springer; 2009; pp. 176-83

[34] Tullis, T., Albert, B.: Measuring the user Experience: Collecting, Analyzing, and Presenting Usability Metrics. Burlington: Elsevier; 2008

[35] Wilson, G. F.: Psychophysiological Test Methods and Procedures. In: Charlton, S. G., O'Brien, T. G., editors: Handbook of Human Factors Testing and Evaluation. 2nd ed. Mahwah: Lawrence Erlbaum Associates; 2002; Chapter 7, pp. 127-56