

Report of first experiments on continuous presence assessment and motivation enhancement

Deliverable D4.2

List of Partners: ETH Zurich, CH (ETH Zurich)
 Hocomo AG, Volketswil, CH (HOCOMA)
 University of Ljubljana, SI (UL)
 Universitat de Barcelona, ES (UB)
 Neurological Clinic Bad Aibling, DE (NKBA)

Document Identifier: MIMICS-D4.2-pu.pdf
Version: 3.0
Due Date: 2009-06-30
Organisation: NKBA
Deliverable: 4.2
Milestone: 4.2
Work Package: 4
Task: 4.1 and 4.2
Dissemination: public
Authors: Matjaz Mihelj, Domen Novak, Marko Munih, Robert Riener, Marc Bolliger, Alexander König, Friedemann Müller, Mark Sapa, Martin Simnacher
Approved by: Lars Lünenburger, Matjaz Mihelj

1	Summary	4
2	Common physiological measurement infrastructure	5
3	Lokomat	7
3.1	Questionnaires	7
3.2	Biomechanical measurements with the Lokomat	8
3.3	Experiments	8
3.3.1	Influence of Lokomat parameters on psycho-physiological parameters	8
3.3.2	Participants	8
3.3.3	Tasks	8
3.3.4	Modelling the influence of physical effort	11
3.3.5	Influence of audiovisual feedback on psycho-physiological status	13
3.3.6	Discussion	18
4	HapticMaster	19
4.1	Questionnaires	19
4.2	Biomechanical measurements	20
4.3	Experiment 1: Initial study	20
4.3.1	Goal	20
4.3.2	Tasks	20
4.3.3	Experimental procedure	21
4.3.4	Participants	22
4.3.5	Results – task performance	22
4.3.6	Results – questionnaires	22
4.3.7	Results – physiological measurements	23
4.3.8	Discussion and conclusions	24
4.4	Experiment 2: Arousal and valence in the presence of physical activity	24
4.4.1	Goal	24
4.4.2	Task	24
4.4.3	Experimental procedure	25
4.4.4	Biomechanical measurements	27
4.4.5	Participants	27
4.4.6	Data analysis methods	27
4.4.7	Results – questionnaires	28
4.4.8	Results – biomechanical measurements	29
4.4.9	Results – physiological measurements	31
4.4.10	Discussion – biomechanical measurements	35
4.4.11	Discussion – psycho-physiological measurements	35
4.4.12	Discussion - Inter-subject variability	36
4.4.13	Conclusions	37
4.5	Experiment 3: Psychophysiology in post-stroke upper limb rehabilitation	37
4.5.1	Goal	37
4.5.2	Tasks	38
4.5.3	Experimental procedure	40
4.5.4	Biomechanical measurements	41
4.5.5	Participants	41
4.5.6	Results	41

5	Conclusion and Outlook.....	45
6	References.....	46

1 Summary

In order to achieve the objectives of the MIMICS project it is necessary to evaluate the impact of the developed technology in neurological patients. The purported increase in motivation of individuals using the systems has to be assessed using both subjective (interview) techniques and quantitative psycho-physiological measurements. These measurements will have to be usable for the feedback loops. A correlation of these measurements to motivation as assessed by using standardized, clinically validated questionnaires has to be demonstrated.

This report summarizes the first experiments to find feasible technological solutions to assess presence, attention and other tractable substrates of motivation on-line during the training.

The general strategy for both devices (with common measurement setup as described in D4.1) was to establish techniques for measuring the psycho-physiological variables of heart rate, skin conductance and skin temperature while controlling either the physical conditions of the subjects (physical load, body weight support, walking speed, etc.) or evoking defined psychological states (challenged, frustrated, bored, nervous) in order to differentiate specific physiological response patterns to the physical and psychological stimuli.

These response patterns were used to estimate the subjects' state regarding separate theoretical constructs such as physical load, valence and arousal. Using a model built from these first experiments, the subjects' physiological reaction to increased physical load was estimated and accounted for in order to gain insight into the current psychological state.

To assess the acuity of this psychological estimate, several well-established psychological self-report instruments (such as the Self Assessment Mannikin and the Behavioral Inhibition System; Behavioral Activation System) were used.

The construct of "presence" as described in the Description of Work proved to be challenging to objectify in first experimental setups, especially once the physiological reaction to increased physical load was taken into account. As this had already been discussed in the Barcelona review, the consortium chose to follow the reviewers' advice to refocus on different psychological constructs that promised to be more tractable in an automated feedback loop.

2 Common physiological measurement infrastructure

The experimental hardware setup and the algorithms for analysis both for HapticMaster at Ljubljana and Lokomat at ETHZ and NKBA were as follows:

The electrocardiogram (ECG) was recorded using pre-gelled, disposable surface electrodes affixed to the chest and abdomen. Skin conductance was measured using a g.GSR sensor (g.tec Medical Engineering GmbH). The electrodes were placed on the medial phalanges of the second and third fingers of the non-dominant hand using Velcro™ straps. The sensor kept the voltage between the two electrodes constant and measured the current between the electrodes in order to estimate skin conductance. This procedure is explained in a classic paper by Fowles et al. (1981). Respiratory rate was obtained using a thermistor-based SleepSense Flow sensor. This sensor is placed beneath the nose and can measure respiration both through the nose and through the mouth. Peripheral skin temperature was measured using a g.TEMP sensor (g.tec Medical Engineering GmbH) attached to the distal phalanx of the fifth finger using medical adhesive tape. These signals were amplified and sampled at 2.4 kHz using a g.USBamp amplifier (g.tec Medical Engineering GmbH) connected to a dedicated signal measurement PC using a USB interface. After the experiment, the signals were band-pass filtered off-line and the relevant parameters were extracted for each time period.

Analysis of the ECG began by extracting the R-peaks (which are the highest peaks in the ECG and correspond to the depolarization of the ventricles during each heartbeat) and manually removing any ectopic beats or noise that had not been removed by filtering. The times between two normal heartbeats (NN intervals) were calculated from the locations of the R-peaks and converted into heart rate. Mean heart rates were calculated for each period of the experiment. After calculating heart rate, several standardized measures (defined in Task Force, 1996) were used to estimate heart rate variability (HRV). In the time domain, the standard deviation of NN intervals (SDNN), the square root of the mean squared differences of successive NN intervals (RMSSD) and the number of interval differences of successive NN intervals greater than 50 ms divided by the total number of NN intervals (pNN50) were calculated. For frequency-domain analysis of HRV, the NN intervals were converted into an instantaneous time series with a constant sampling frequency using cubic spline interpolation and the power spectral density (PSD) of this time series was calculated using Welch's method of modified periodograms. The PSD has two frequency bands of interest to us: the low-frequency band (LF) between 0.04 Hz and 0.15 Hz and the high-frequency band (HF) between 0.15 Hz and 0.4 Hz. Three frequency-domain estimates of HRV were calculated: total power in the LF band, total power in the HF band (commonly referred to as respiratory sinus arrhythmia – RSA) and the ratio of the two (commonly referred to as the LF/HF ratio).

The skin conductance signal can be divided into two components: the skin conductance level (SCL) and skin conductance responses (SCRs). The SCL is the baseline level of skin conductance in the absence of any particular discrete

environmental event. Its mean value was calculated for each time period. The absolute value of skin conductance could not be measured with our instrument, which records only changes from an initial offset, so the value of skin conductance at the beginning of the experiment was considered to be the zero value. SCRs are temporary increases in skin conductance followed by a return to the tonic level. They can occur in response to strong stimuli, but also occur in the absence of any specific event, even when the subject is resting. Every increase in skin conductance was classified as a SCR if its amplitude exceeded $0.05 \mu\text{S}$ and the peak occurred less than five seconds after the beginning of the increase. SCR frequency for each period was calculated by dividing the total number of detected SCRs with the length of the period (which was always five minutes).

Mean respiratory rate was calculated in breaths per minute for each time period. The signal obtained from the respiration sensor was a signal which increased during inspiration and decreased during expiration. Respiratory period (the time between two breaths) was calculated by measuring distances between two consecutive peaks in the respiration signal. Respiratory rate was calculated from respiratory period. Additionally, respiratory rate variability was estimated by calculating the standard deviation of the respiratory rate time series.

Peripheral skin temperature was recorded at the end of each time period by averaging temperature during the last five seconds to remove the effect of noise.

3 Lokomat

Two main experiments have been performed at ETH Zurich. In the first experiment we evaluated the influence of the Lokomat parameters walking speed, body weight support and guidance force on psycho-physiological parameters from subjects walking in the device. The goal of this study was to determine the influence of different parameters of standard Lokomat training on psycho-physiological parameters to be able to differentiate between physical and psychological causes of changes in the recorded signals. In the second experiment different virtual scenarios were created that aim to induce three different psychological states (bored, nervous and pleasant) in the subjects walking in the Lokomat. The goal of this study was to investigate whether it will be possible to differentiate between different psychological states just by investigation of the physiological recordings.

All measurements for Deliverable 4.2 were conducted with the downscaled Lokomat setup from Deliverable 4.1 (Figure 3-1). Psycho-physiology was assessed with the parameters galvanic skin response (GSR), electrocardiogram (ECG), peripheral skin temperature and breathing frequency. Furthermore, we recorded video and Lokomat forces.



Figure 3-1: Downscaled Lokomat set-up used for experiments in D4.2

3.1 Questionnaires

Emotions were assessed in the context of an adapted arousal-valence emotion model from Russel (1980). Therefore the established Self-Assessment Manikin (SAM) (Lang, 1980) was used for self-report questionnaires. The SAM is a non-verbal pictorial assessment technique that directly measures the pleasure, arousal, and dominance associated with a person's affective reaction to a wide variety of stimuli. In Figure 3-2 the SAM for assessing valence is displayed.

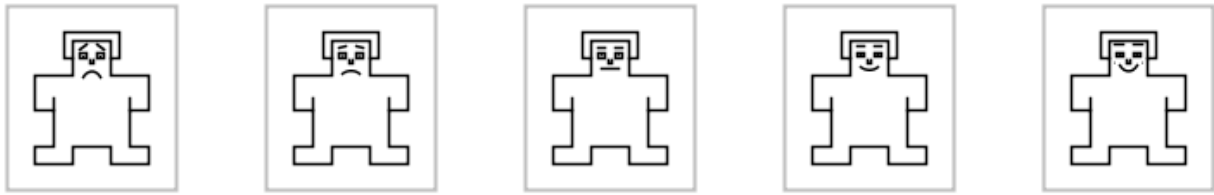


Figure 3-2: Self-Assessment Manikin to assess arousal. The figures are presented to the volunteers who have to estimate their state.

Additionally involvement of subjects in virtual environment and their immersive tendency was assessed by a questionnaire from Witmer and Singer (1998). The questionnaires for assessing involvement had to be slightly adapted to the Lokomat setup.

3.2 Biomechanical measurements with the Lokomat

The Lokomat allows to measure forces acting between the subjects and the robotic device. These forces can be used to estimate the physical work of subjects walking in the Lokomat.

3.3 Experiments

Psycho-physiological reactions are influenced by physical and psychological stimuli. Therefore, to be able to control psycho-physiological parameters the influence of both stimuli alone on these parameters has to be known. Thus, two different studies have been prepared. In the first study the influence of standard Lokomat training on psycho-physiological reactions was examined without additional input from the virtual reality (VR) system. In a second study influence of audiovisual stimuli on psycho-physiological parameters in a standard Lokomat training was assessed. The results from these two studies are reported in the following sections.

3.3.1 Influence of Lokomat parameters on psycho-physiological parameters

In this study Lokomat parameters body weight support, guidance force and treadmill speed were changed and psycho-physiological parameters from subjects walking in the device were recorded. The goal of this study was to determine the influence of different parameters of standard Lokomat training on psycho-physiological parameters to be able to differentiate between physical and psychological causes of changes in the recorded signals.

3.3.2 Participants

All measurements were performed with healthy volunteers without any neurological disease.

3.3.3 Tasks

Influence of body weight support (BWS): Subjects (n=5) had to walk with four different levels of BWS (0%, 30%, 50%, 70%) at a constant treadmill speed of 3 km/h. The order of BWS conditions was randomized. Each condition was recorded for 5 minutes. ECG, GSR, Skin temperature and breathing frequency were recorded. Results showed that none of the recorded parameters were affected by BWS in

healthy young volunteers. Figure 3-3 shows the results for the changes in heart rate over the 4 different BWS conditions.

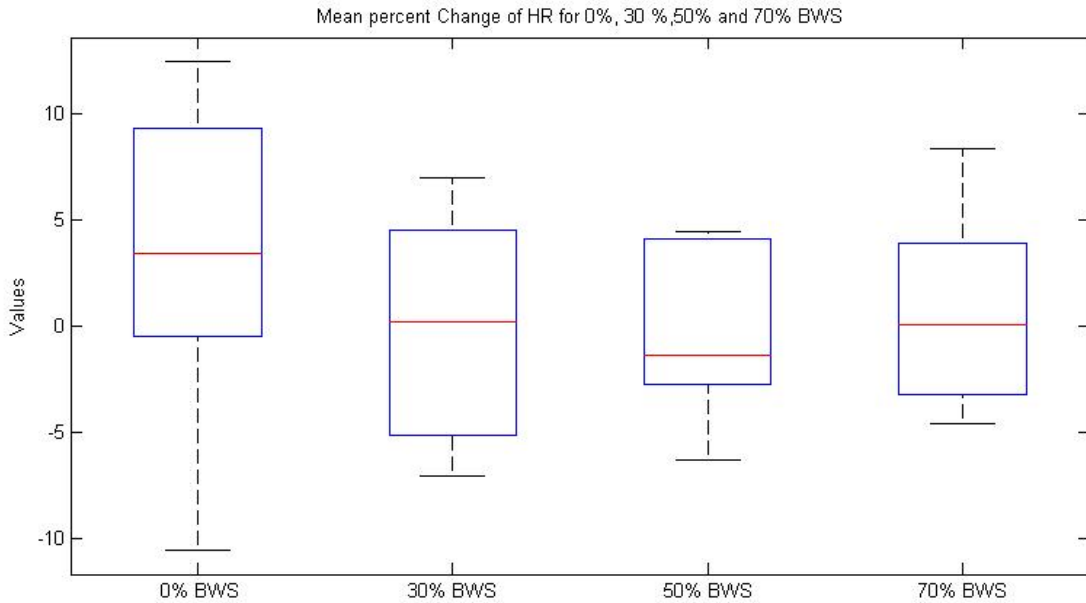


Figure 3-3: Mean percentage change of heart rate during walking with different body weight support.

Influence of treadmill speed: Subjects (n=5) had to walk with two different treadmill speeds (2.0 and 3.0 km/h) without the Lokomat, within the Lokomat with passive behaviour and with Lokomat with active behaviour. The three conditions were randomized. Results showed that ECG (and the derived parameters) depends significantly ($p < 0.05$) on the work load subjects have to perform (passive – active behaviour, Figure 3-4). The influence of treadmill speed showed only a tendency on ECG (increasing speed – increasing ECG) but was not significant. All other parameters showed no significant ($p > 0.05$) changes.

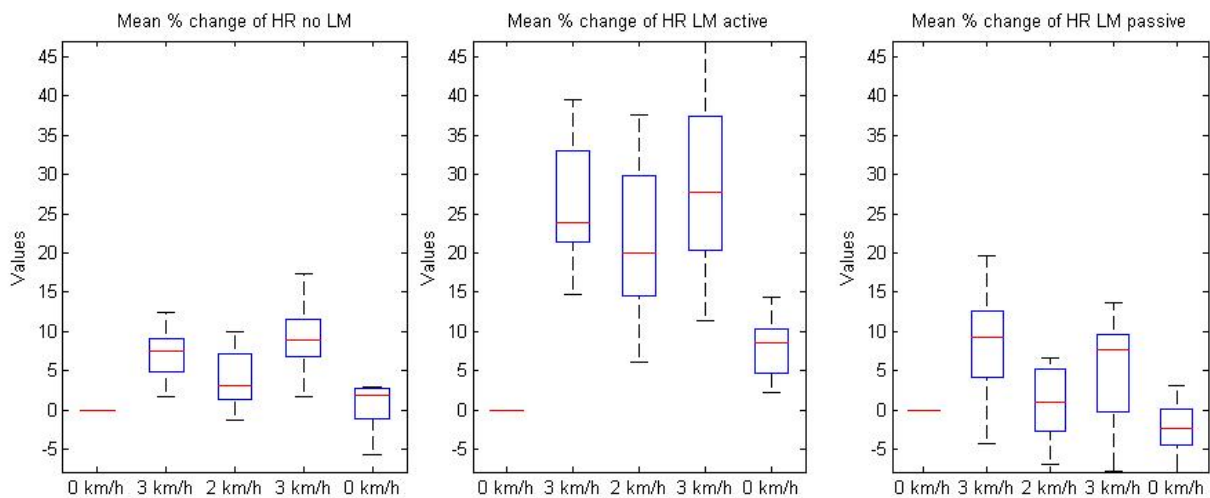


Figure 3-4: Influence of treadmill speed in three different conditions on heart rate. HR was significant increased in the active Lokomat conditions.

Influence of guidance force (GF): Subjects (n=5) had to walk with three different GF (0%, 50% and 100%) at two different speeds (2 km/h and 3 km/h). Guidance force is a training parameter that controls the amount of support of the Lokomat. The GF conditions were randomized. Results showed that GF has no significant influence on psycho-physiology in healthy volunteers. Figure 3-5 shows the results for breathing frequency changes.

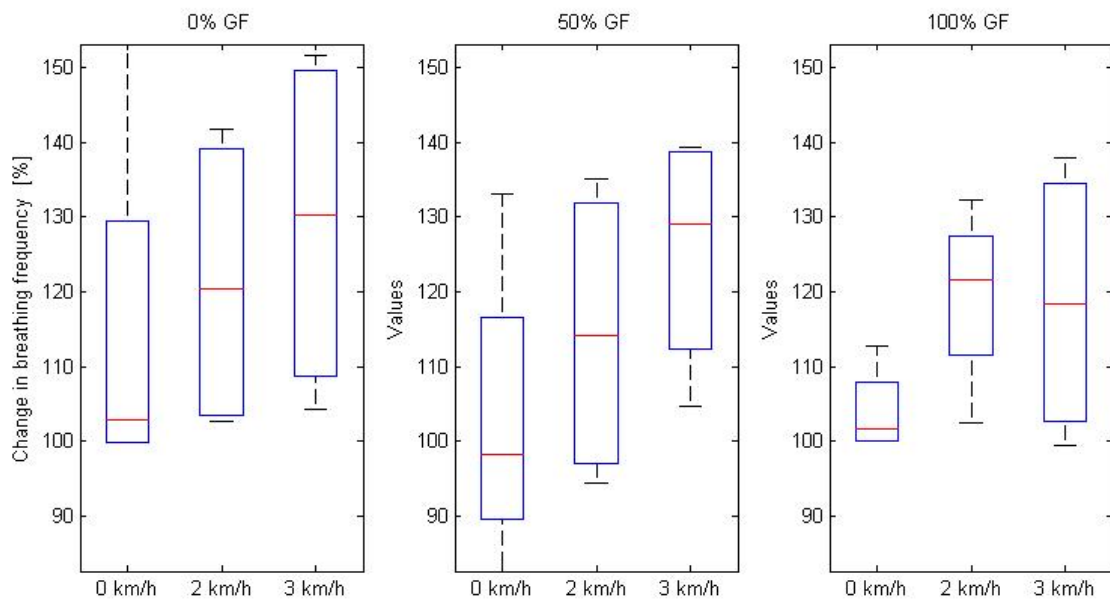


Figure 3-5: Change in breathing frequency due to walking with different guidance forces at 3 different speeds. There was a tendency that with higher walking speed breathing frequency increased.

Tests with patients

Further tests with subjects with neurological disorders will be performed at Bad Aibling in July and August 2009. We expect BWS and GF to also have an effect on these subjects.

3.3.4 Modelling the influence of physical effort

As shown above, the Lokomat has a non-negligible influence on the subject's physiology, mainly due to the increased physical effort of walking compared to standing. This influence cannot be modelled for skin conductance, skin temperature or breathing frequency, as the underlying neurological and physiological mechanisms are too complex.

The effect of changes in treadmill speed, body weight support or guidance force on the heart rate have been modelled. Any increase in physical effort will increase the mean heart rate of a subject. We used the above given data to model this effect and identified a non-linear model with eleven parameters. For our subject group, seven of these could be assumed to be subject-independent, while four parameters were subject-specific.

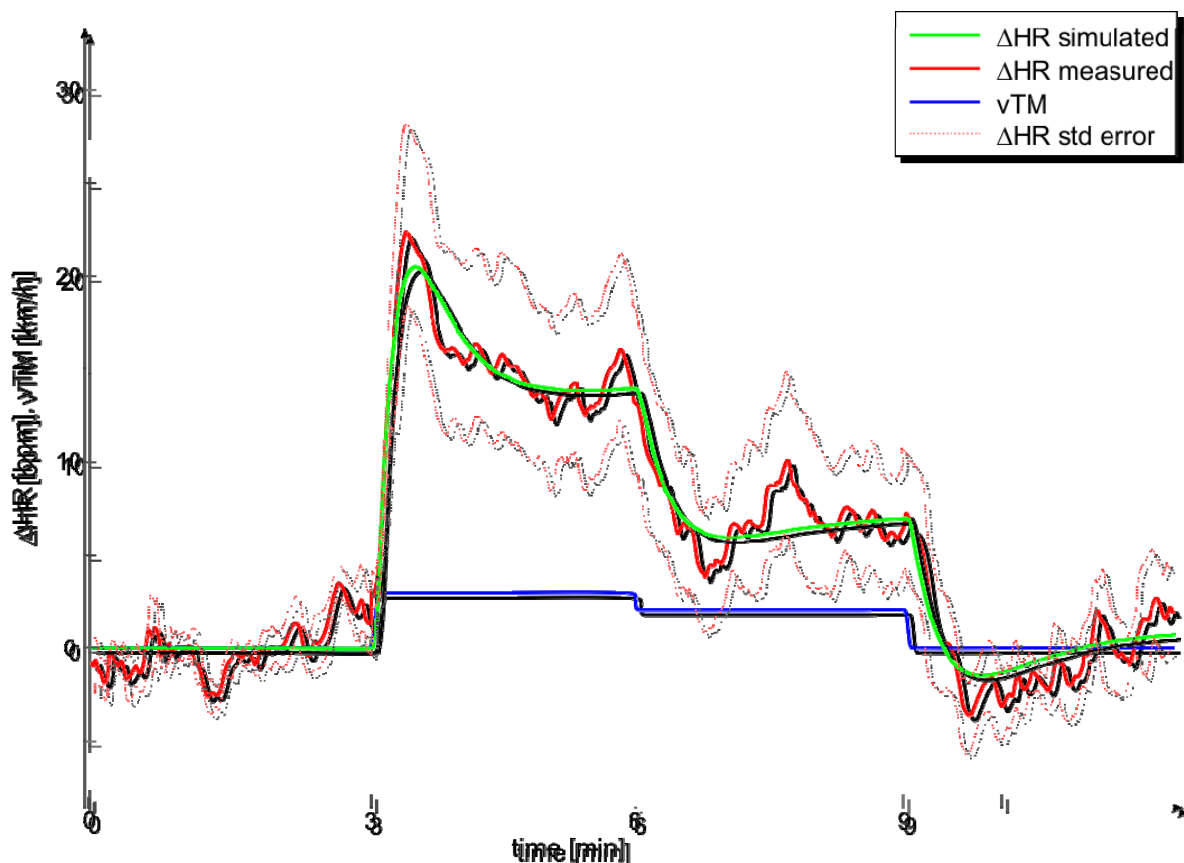


Figure 3-6: Heart rate as a function of treadmill speed and its prediction from our model. Green line: simulated HR, red line: measured HR, blue line: treadmill speed, red dotted line: +/- std error.

Using this model, we can predict the change in heart rate caused by the physical effort performed by the subject in the Lokomat. Table 3-1 summarizes the prediction quality we were able to reach for five subjects using the walking speed profile given in Figure 3-6. Subject 2 did not react to the intervention with an increase of heart rate. Therefore, the prediction quality of the heart rate was around $R^2=0.3$.

Future experiments will show whether we can subtract the contribution of the physical effort to the change in heart rate and identify the contribution of psychological effects evoked by the multimodal display. These studies will also be undertaken with patients.

Table 3-1: Results from predicting the heart rate of healthy subjects.

Subject	R^2 for optimization of eleven parameters	R^2 for optimization of four Parameters
1	0.90	0.90
2	0.32	0.30
3	0.93	0.88
4	0.93	0.92
5	0.94	0.94

R^2 : coefficient of determination ($R^2=1-SS_{err}/SS_{tot}$, SS_{err} : the sum of squared errors, SS_{tot} : the total sum of squares)

3.3.5 Influence of audiovisual feedback on psycho-physiological status

During Lokomat training, the virtual environment can have an influence on the patient on two different levels: on the one hand, the subject can be influenced by the multimodal stimulation (graphics, haptics, sound) itself. On the other hand, the task that needs to be solved can influence the psychological state of the subject. A task that is too difficult can produce stress while a task that is too easy can bore the subject.

The goal of this study was to investigate whether it will be possible to differentiate between different psychological states just by investigation of the physiological recordings.

In study I, we created different virtual scenarios that aim to induce three different psychological states in the subjects: bored, nervous and pleasant (Figure 3-7). Subjects had to walk in the Lokomat while different VR scenarios were presented to them. Two main scenarios were developed: a city scenario and a canyon scenario (Figure 3-8).

In the city scenario subjects walked on a sidewalk through a city. On each side of the sidewalk a street was placed. Details can be seen in Figure 3-8. In the city scenario the following parameters could be changed: traffic density, atmosphere, weather, brightness and sound. The three city scenarios were adapted as following to induce the three states bored, pleasant and nervous:

- *City bored*: no sound, no traffic, broad sidewalk, neutral atmosphere
- *City pleasant*: relaxing music (Buena Vista Social Club: Chan Chan), moderate traffic, sunset (atmosphere).
- *City nervous*: scary music, heavy traffic, narrow sidewalk, strong rain, thunderbolt, dark atmosphere

In the canyon scenario subjects walked on a bridge over a river. Details can be seen in Figure 3-8. In the canyon scenario the following parameters could be changed: atmosphere, weather, brightness, sound, depth of water (river), quiet or white-water, width of bridge, handrails visible/hidden. The three canyon scenarios were adapted as following to induce the three states bored, pleasant and nervous:

- *Canyon bored*: no sound, broad bridge, handrails, neutral atmosphere
- *Canyon pleasant*: relaxing music (Buena Vista Social Club: Chan Chan), light water noise, handrails, sunset (atmosphere)
- *Canyon nervous*: scary music, white-water, loud water noise, narrow bridge, no handrails, strong rain, thunderbolt, dark atmosphere

To increase the effect of the VR sound effects were implemented in the scenarios. Questionnaires (Self-Assessment Manikin, SAM) were used to verify whether subjects experienced the different psychological states.

In study II, we investigate how a virtual task will influence the subjects and if we are able to induce states, in which the subject feels under-challenged, over-challenged or “optimally” challenged.

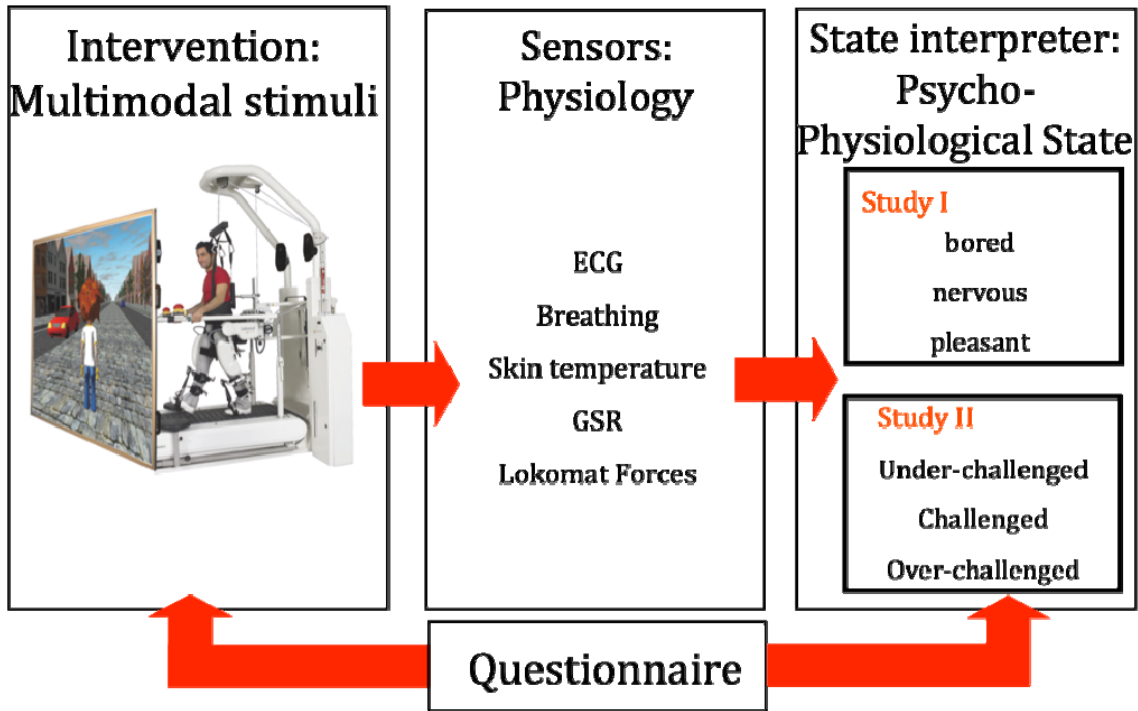
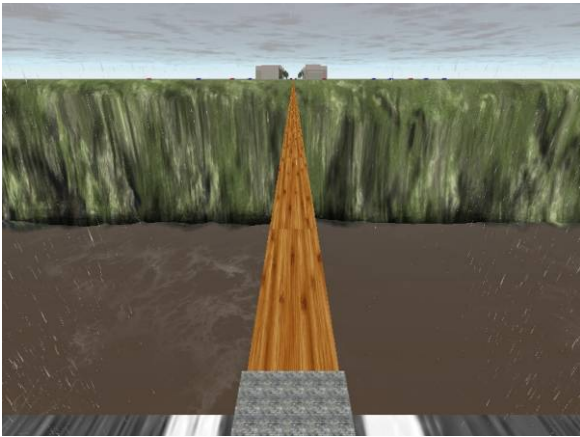


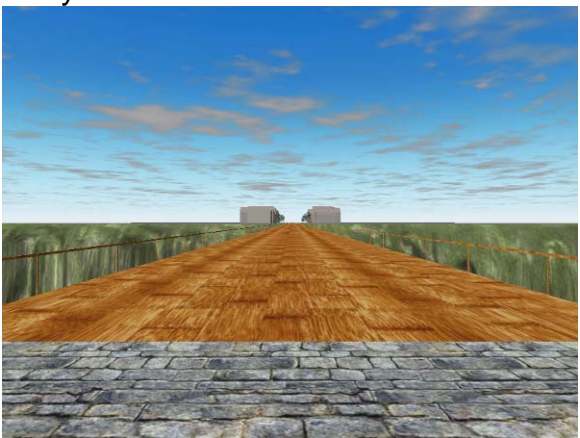
Figure 3-7: Setup for studying influence of virtual reality on psycho-physiology



Canyon nervous



City nervous



Canyon bored



City bored



Canyon pleasant



City pleasant

Figure 3-8: Canyon and city scenarios to induce nervous, bored and pleasant state

Study I: First subjects had to fill out questionnaires (presence and immersive tendency). Measurement started with a 5 minutes standing period followed by a baseline measurement (5 min) while they had to walk within the Lokomat without VR. Afterwards subjects had to walk within the Lokomat while the 6 different VR scenarios (City nervous (N-City), city bored (B-City), city pleasant (P-City), canyon nervous (N-Can), canyon bored (B-Can), canyon pleasant (P-Can)) were presented in randomized order. Between each two VR scenarios SAM was assessed. The scenarios were followed by an additional baseline and standing measurement (Figure 3-9). 10 healthy volunteers were measured for study I.

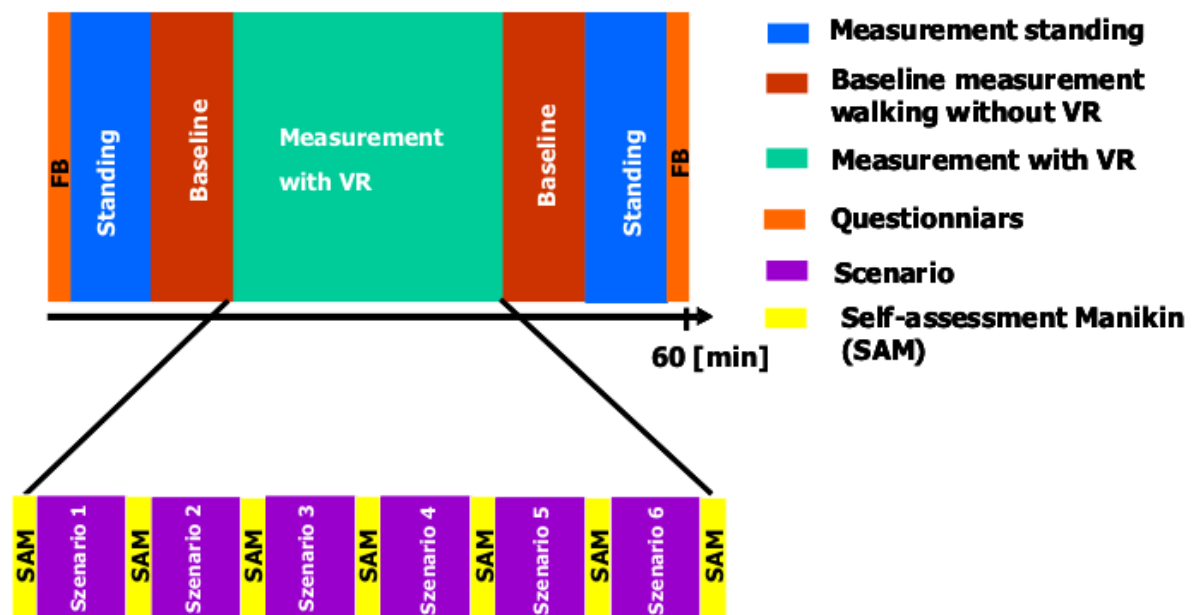


Figure 3-9: Measurement protocol for study I.

Results showed that neither psycho-physiological parameters nor Lokomat forces of study I were significantly affected by any of the VR scenarios. Figure 3-10 shows the results for the changes of heart rate due to the 6 different VR scenarios. However, questionnaires revealed that subjects experienced the desired states. Arousal (Figure 3-11) and valence differed significantly between the scenarios while physical effort was not affected. Results in questionnaires about involvement of subjects in VR showed that subjects tolerated VR and felt present in the scenarios.

In summary we can say that the states induced by the stimuli used could not be distinguished based on the psycho-physiological parameters when using the described VR scenarios. We concluded that stronger stimuli in VR or the implementation of specific tasks in the scenarios are needed in order to induce stronger psycho-physiological states. Therefore we implemented tasks (collecting items while walking in VR) in the same city scenario for study II. This study is currently running.

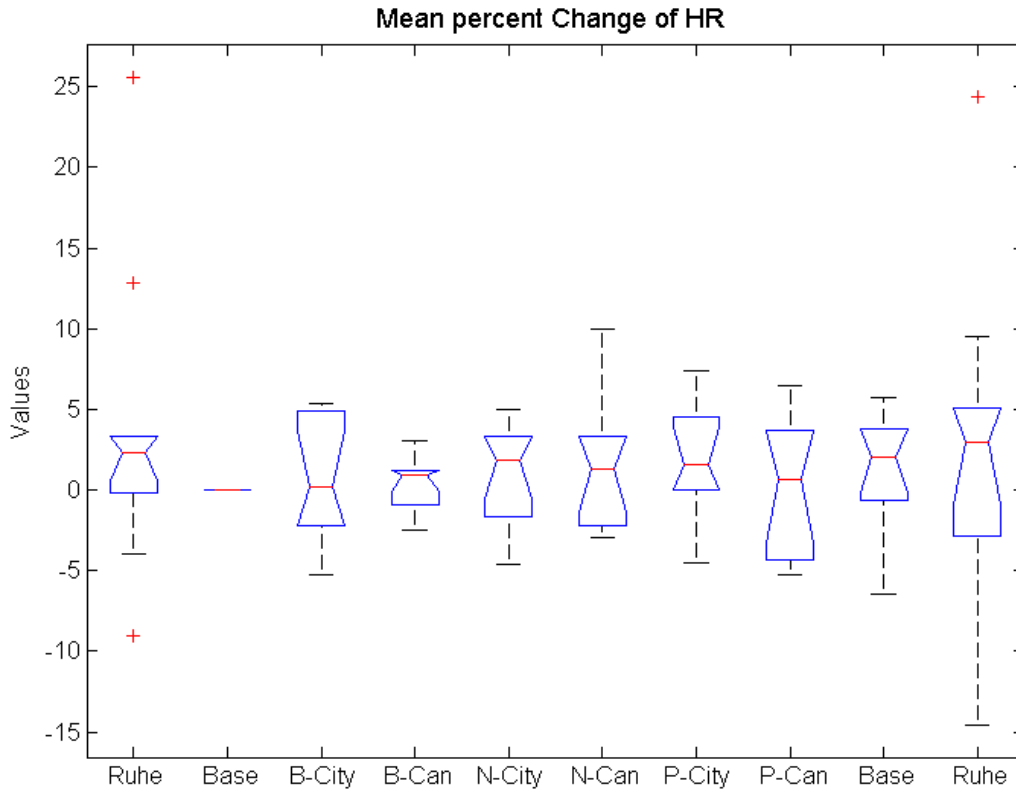


Figure 3-10: Changes of heart rate over the 6 different scenarios

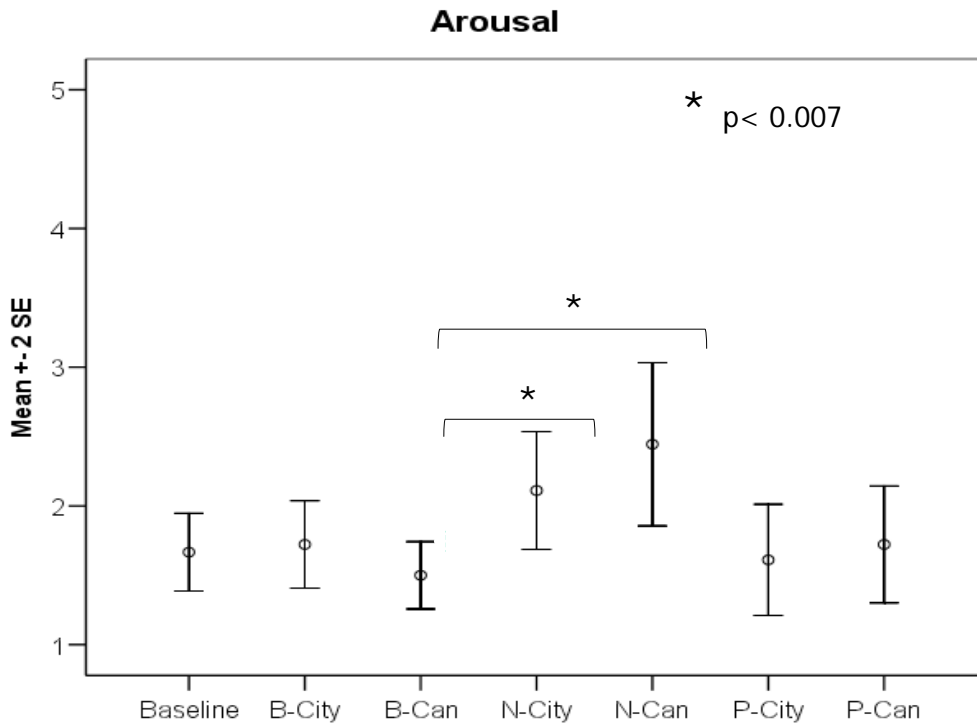


Figure 3-11: Results of self-assessment Manikin questionnaires on arousal. Can: canyon, B: bored, N: nervous, P: pleasant

3.3.6 Discussion

Moderate stimuli to change the psychological state of subjects walking in the Lokomat did not evoke significant changes in the measured psycho-physiological parameters. None of the signals (GSR, ECG, Breathing frequency and skin temperature) did show significant changes due to different VR scenarios. However, questionnaires revealed the subjects experienced different psychological states while different VR scenarios were presented to them. We conclude that the physical activity induced by walking in the Lokomat has a much stronger influence on the signals than the moderate stimuli induced by different VR scenarios. In a next step we will now increase the stimuli to affect the psychological state of our subjects. We expect the stronger stimuli will induce stronger changes in the psychological states and that this will be reflected in significant changes in the psycho-physiological parameters. This will be done by implementing tasks in the VR scenarios. As we already know from former experiments we have to keep the physical activity of the subjects as constant as possible while they perform the different tasks. This will be controlled by assessing the forces acting between the Lokomat and the subject.

Additionally, we saw that all assessed psycho-physiological parameters showed a great variability between the subjects. This goes in line with the results from the experiments assessed at UL.

All the points listed above have to be taken into account before we can make any suggestions for a decision making system.

4 HapticMaster

Three consecutive experiments have been performed at UL. The first was an initial experiment and primarily served to determine whether we could show differences between tasks or between difficulty levels. The other two were performed using the upper extremity system as described in deliverable D1.2. These experiments aimed to show that psycho-physiological responses provide significant data about psychological activity even in the presence of physical activity (which also affects psycho-physiological responses) and that they can be used with stroke patients in virtual-reality-assisted rehabilitation settings.

4.1 Questionnaires

For the very first experiment, participants were asked to rate their level of satisfaction, frustration and concentration during each task on a six-point scale, with zero representing “did not feel at all”, one representing “very low”, three representing “medium”, and five representing “very high”. They were also asked to rate the difficulty of the task on a five-point scale from “very low” to “very high”. No questionnaires were presented during rest periods.

After the first experiment, we decided to examine emotions within the context of the arousal-valence emotion model (Russell, 1980). In this model, a person’s psychological state is described as a combination of arousal, which represents a general level of mental activity, and valence, which indicates whether a person’s emotions are positive or negative. Thus, for the following experiments, nine-point arousal and valence scales from the self-assessment manikin (Lang, 1980) were used as the primary self-report questionnaires. These scales allow the participant to rate their level of general mental arousal and emotional valence graphically by choosing the picture they feel best represents their current mood. The questionnaire was shown to each participant and the meaning of each scale was explained before the experiment began. The questionnaire was then presented after each period of the experiment. The self-assessment manikin was chosen because of its simplicity, requiring only very brief breaks between the task periods.

Following the second experiment, it became apparent that the behaviour and physiology of the participants were dependent not only on how they felt in regard to arousal and valence, but also on how they dealt with challenges. Some responded to challenging situations by becoming upset and giving up while others tried even harder. Thus, for the third and fourth experiment, we presented participants with the BIS/BAS (Behavioral Inhibition System / Behavioral Activation System) scales (Carver & White, 1994). A popular psychological tool, the BIS/BAS questionnaire consists of 24 statements that participants can agree or disagree with. The answers to these questions are graded on one BIS (Behavioral Inhibition System) subscale and three BAS (Behavioral Activation System) subscales: Reward Responsiveness, Drive and Fun-seeking. The BIS measures a tendency towards negative emotional responses and anxiety, while the BAS measures motivation and movement toward goals, seeking for fun and seeking for pleasurable situations. The three BAS scales focus on different aspects of incentive sensitivity. In particular, Fun Seeking is known to have elements of impulsiveness that are not contained in the other scales.

4.2 Biomechanical measurements

The HapticMaster allows real-time three-dimensional measurement of the position of the end-effector as well as the forces and torques acting on the end-effector. In addition, it can measure the force with which the user grips a virtual object. However, since the different tasks are different in nature, we cannot calculate the same biomechanical parameters in every task. Of course, in tasks that do not include the HapticMaster, no biomechanical measurements can be taken at all. Thus, the biomechanical parameters measured will be presented for each task separately.

4.3 Experiment 1: Initial study

4.3.1 Goal

The purpose of the first experiment was simply to examine the physiological responses of participants in response to simple tasks. First of all, we wished to test the equipment and see if there would be a physiological difference between rest and task periods. Our secondary objectives were to see whether there were physiological differences between different task difficulties or task types.

4.3.2 Tasks

Two different tasks were used for the experiment: a mental arithmetic task and a manual error correction task. Each participant performed each task by itself (single-task period) and both tasks at the same time (dual-task period).

The mental arithmetic task presented the participants with two numbers that had to be multiplied. These numbers were randomly generated between zero and thirty for each subject. Four different possible answers were shown immediately underneath. One of the answers was correct while the other three were generated by adding or subtracting a random multiple of ten (between -40 and +40) to the correct answer. Using speech recognition, the participants verbally chose the answer they believed was correct (by saying "first", "second", "third" or "fourth"). If the participants answered correctly, their choice was coloured green. If the participant answered incorrectly, their choice was coloured red and the correct choice was coloured green. The participants had fifteen seconds to answer each question; if they failed to answer within that time, the result was identical to making an incorrect choice (except that no number turned red). The time remaining was displayed using a large bar next to the numbers which grew progressively shorter and turned from green through yellow to red as time ran out. After a choice was made or time ran out, there was a five-second pause followed by the next two numbers to be multiplied.

The manual error correction task presented the participants with a virtual inverted pendulum. A thin pole with a weight at its top end is attached at its bottom to a moving cart. This vertical pendulum is inherently unstable and must be actively balanced by moving the cart horizontally. The participants were presented with a visualization of a simulated cart and pole on the computer screen and moved the cart using a haptic interface. The cart moved in the same direction and with the same velocity as the end of the haptic interface. The model dynamics were adjusted in such a way to make balancing the pendulum moderately challenging. If the participants failed to balance the pendulum and it fell to a horizontal position, it was

immediately reset to a nearly vertical position. Force feedback was also implemented with the haptic interface, allowing the participants to feel the reaction forces resulting from the movement of the cart. The movement of the endpoint was not restricted, allowing participants to choose the most comfortable position of the haptic interface. However, only movement left and right affected the virtual cart. During the dual-task period, both tasks had to be performed simultaneously and were shown on the same 22-inch display, one above the other (Figure 4-1). During a single-task period, the task not being performed was replaced by a uniform gray background. For the haptic interface, we used a Phantom Premium from SensAble Technologies, Inc. This device provided a range of motion approximating hand movement pivoting at the wrist. It included a passive stylus and thimble gimbal and provided three degrees of freedom force feedback and three degrees of freedom positional sensing. Speech was recorded using a headset.

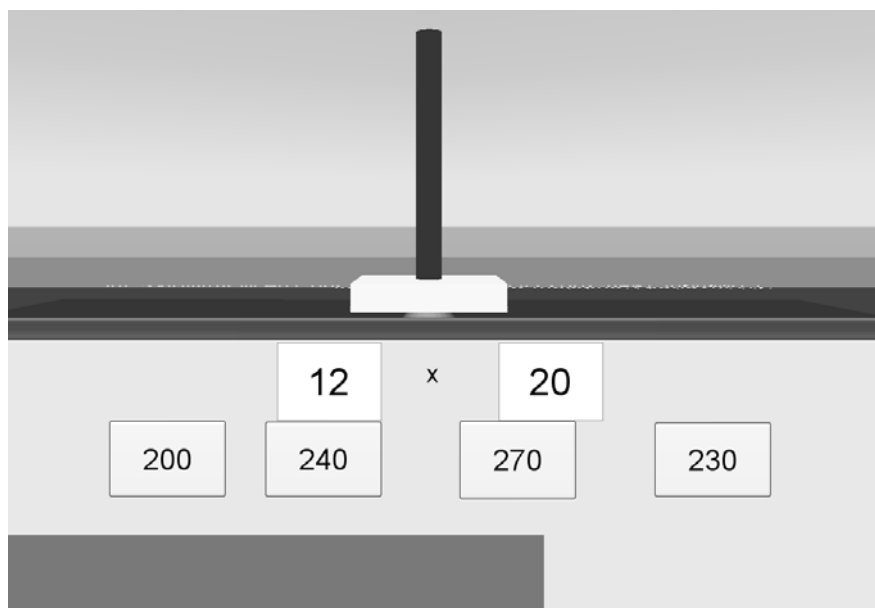


Figure 4-1 Computer display for the dual-task situation

Manual error correction task performance was evaluated by calculating the mean time between pendulum resets and the standard deviation of the time between pendulum resets. Mental arithmetic task performance was evaluated by calculating the percentage of correct answers, the mean time needed to answer a question and the standard deviation of the time needed to answer a question.

4.3.3 Experimental procedure

The experiment was conducted in a quiet area of the laboratory where external stimuli did not disturb the participants. The temperature and humidity in the laboratory were kept constant. There was never more than one participant inside the laboratory at any time. Before their arrival, each participant was randomly assigned the order in which they would perform the tasks. Since there were three task periods and thus six possible orders, four participants were assigned to each. Upon arrival in the laboratory, the purpose of the experiment and both tasks were described and the participant was allowed to practice each task as well as both tasks simultaneously. None chose to practice either a single-task or dual-task situation longer than five

minutes. The measurement equipment was then attached and turned on. The signals were monitored until the participant became used to the environment and most effects of nervousness or novelty had disappeared. Then, the three task periods (two single-task and one dual-task) were conducted in random order. Each task period lasted five minutes and was preceded by a five-minute rest period during which the participant rested quietly and baseline values of psychophysiological parameters were obtained. After each task period, the participant was presented with a self-evaluation questionnaire (detailed in section 4.2). Between the rest period and the manual error correction task, participants were asked to move the Phantom left and right with the computer display turned off for five minutes. They were specifically asked to move as if they were performing the manual error correction task that they had previously practiced. This part of the experiment, hereafter referred to as the movement period, was necessary in order to gauge the effect that physical effort alone had on physiological responses. After completion of the tasks, the equipment was switched off and an informal interview was conducted about the entire experience. The entire experiment took about 70 minutes.

4.3.4 Participants

Twenty-four male students and staff members of the Faculty of Electrical Engineering in Ljubljana participated in the experiment. All were healthy, without any major cognitive or physical deficits. They were aged between 20 and 46, mean age 28.0 years, median age 26 years, standard deviation 6.6 years. Each subject signed an informed consent form after the purpose and procedure of the experiment was explained to him.

4.3.5 Results – task performance

During the mental arithmetic single-task period, participants correctly answered 80.4% of all questions. The time needed to answer a question was 7.7 +/- 3.9 seconds (mean +/- standard deviation). During the dual-task period, participants correctly answered 73.3% of all questions (fewer than during the single-task period, $p = 0.03$). The time needed to answer a question was 8.5 +/- 4.8 seconds. There was a significant difference in both mean answer time ($p = 0.006$) and standard deviation of answer time ($p = 0.02$) between the single-task and dual-task periods.

Without participant input, the inverted pendulum fell to the ground and was reset every 6.0 seconds. When balanced by participants during the manual error correction single-task period, it fell and was reset every 19.2 +/- 5.4 seconds. During the dual-task period, the pendulum fell and was reset every 13.5 +/- 7.6 seconds. There was a significant difference in both mean reset time ($p = 0.004$) and standard deviation of reset time ($p = 0.01$) between the single-task and dual-task periods.

4.3.6 Results – questionnaires

Data from the questionnaires is summarized in Table 4-1. There was no significant difference between the two single-task periods for any of the four reported subjective feelings. The dual-task period resulted in the highest subjective feelings of frustration, concentration and difficulty together with the lowest satisfaction. The differences in frustration and task difficulty between the dual-task period and either single-task

period were statistically significant ($p < 0.01$ for either variable when compared to either single-task period).

Correlations between the different variables were analyzed using Pearson correlation. Concentration was significantly correlated with both task difficulty (correlation coefficient 0.505 $p < 0.001$) and with frustration (correlation coefficient 0.399, $p = 0.002$).

Table 4-1 Results of self-report questionnaires.

	manual error correction		mental arithmetic		dual-task	
	mean	st. dev.	mean	st. dev.	mean	st.dev.
satisfaction	2.5	1.3	2.8	1.1	1.9	1.3
frustration	2.0	1.3	1.8	1.2	2.9	1.3
concentration	3.6	0.9	3.8	1.1	4.2	0.7
difficulty	3.2	1.0	3.3	0.5	4.5	0.7

4.3.7 Results – physiological measurements

Table 4-2 shows the differences in physiological responses between baseline (rest) and task as well as between different task periods. The only exception to this is the SCL, which is already measured as deviations from an initial value and is therefore presented as the absolute difference between periods.

Table 4-2 Physiological differences between different task periods. All differences other than mean SCL are presented as percentage of baseline (rest) value. Bold underlined values and asterisks indicate statistical significance: * for $p < 0.05$, ** for $p < 0.01$ and * for $p < 0.001$. Top row abbreviations: B = baseline, M = movement period, MEC = single-task manual error correction, MA = single-task mental arithmetic, D = dual-task period.**

	M - B	MEC - B	MA - B	D - B	MEC - M	MA - MEC	D - MEC	D - MA
mean SCR frequency	13.1	<u>122.7***</u>	<u>165.2***</u>	<u>227.3***</u>	<u>109.6**</u>	42.5	104.6	62.1
mean SCR amplitude	3.8	<u>22.9*</u>	<u>34.9***</u>	<u>46.1*</u>	<u>19.1*</u>	12.0	23.2	11.2
mean SCL (μ S)	0.0	<u>7.9**</u>	<u>12.6***</u>	<u>11.2***</u>	<u>7.9*</u>	<u>4.7*</u>	<u>3.3*</u>	-1.4
mean respiratory rate	<u>16.6***</u>	<u>21.8***</u>	<u>17.9***</u>	<u>16.8***</u>	5.2	-3.9	-5.0	-1.1
respiratory rate variance	<u>-24.0*</u>	<u>-39.1***</u>	22.7	2.9	-15.1	<u>61.8**</u>	<u>42.0***</u>	-19.8
mean HR	<u>2.6*</u>	1.5	<u>2.4*</u>	<u>3.8***</u>	-1.1	0.9	2.3	1.4
SDNN	-4.0	<u>-12.0*</u>	13.6	5.5	-8.0	<u>25.6**</u>	<u>17.5**</u>	-8.1
RMSSD	-9.6	0.0	10.5	10.2	9.6	10.5	10.2	-0.3
pNN50	-9.0	7.3	72.6	24.7	16.3	65.3	17.4	-47.9
LF/HF ratio	39.3	-9.2	33.5	35.2	<u>-48.5*</u>	42.7	44.4	1.7
HF power (RSA)	-12.8	19.5	<u>42.4*</u>	<u>33.3*</u>	<u>32.3*</u>	22.9	13.8	-9.1
LF power	7.1	<u>-4.6*</u>	79.9	62.8	<u>-11.7*</u>	<u>84.5**</u>	<u>67.4*</u>	-17.1
final temperature	0.8	-0.7	<u>-2.2***</u>	<u>-2.7***</u>	<u>-1.5*</u>	<u>-1.5*</u>	<u>-2.0*</u>	-0.5

4.3.8 Discussion and conclusions

The initial study showed several physiological changes in response to both physical and purely mental activity. The three skin conductance parameters and the final skin temperature appeared especially promising, as they were not triggered by pure physical activity (movement period), but only by the tasks requiring mental effort. Heart rate and respiration were significantly affected by physical activity, so their usefulness was uncertain.

However, the study also highlighted a problem: though the mental arithmetic task and the manual error correction task showed no significant differences in responses to the self-report questionnaires, there were distinct physiological differences between the two. Moreover, though there were significant differences in self-report questionnaire responses between single-task and dual-task periods, there was no significant physiological difference between the mental arithmetic single-task period and the dual-task period. It was apparent that psycho-physiological measurements do not necessarily reflect the same thing as self-report questionnaires do.

4.4 Experiment 2: Arousal and valence in the presence of physical activity

4.4.1 Goal

Following the first experiment, we chose to adopt the arousal-valence emotion model (detailed in section 4.2) for further studies. Since the human body has been shown to respond differently to arousal and valence, we decided to plan an experiment in which we could observe the effects of arousal and valence partially independently. As this has not been previously done in a task that requires significant physical load, we planned an experiment where we could study the effects of three psycho-physiological “dimensions”: arousal, valence, and physical load.

4.4.2 Task

Participants were presented with the inverted pendulum task already used in Experiment 1 (section 4.4). For this experiment, the HapticMaster was used for the haptic interface instead of the Phantom. Six different versions of the task were programmed. In the original (neutral) version, the model dynamics were balanced in such a way as to make balancing the pendulum moderately challenging. A second version was made to be difficult to the point of frustration. This was done by introducing a half-second delay between the time the cart was moved and the time the cart’s movement actually affected the pole. Additionally, the gravity acting on the pole was stronger and the pole was less responsive to the movement of the cart. A third version was made to present a lesser challenge. This was done by making the pole more responsive to the movement of the cart and by weakening the gravity acting on the pole. The other three versions of the task were identical to the first three, but more physical load was required to move the HapticMaster. Though the cart and pole had the same virtual mass, the force applied by the participant was divided by five. This forced the participant to apply five times the force that had been applied in the original task for the same movements. Henceforth, these versions of the task will be referred to as the task with high physical load, while the original versions of the task will be referred to as the task with low physical load.

In addition to the inverted pendulum task, a control task was introduced to evaluate the effect of physical load in the absence of mental load. During the control task, participants moved the HapticMaster left and right at an even, moderate speed while nothing was shown on the display and all force feedback was disabled. Two different versions of the control task were made: one with low and one with high physical load. Once again, high physical load was implemented by dividing all the measured forces by five. Before the beginning of the experiment, participants were verbally encouraged to move the HapticMaster at speeds similar to those during the inverted pendulum task.

4.4.3 Experimental procedure

The experiment was conducted in a quiet area of the laboratory where external stimuli did not disturb the participants. In addition, the temperature and humidity in the laboratory were kept constant so as not to influence measurements of electrodermal activity or skin temperature. There was never more than one participant inside the laboratory at any time. Each participant performed the inverted pendulum task in two separate time blocks. Each block consisted of:

- an initial rest period in order to obtain baseline measurements,
- three inverted pendulum task periods, each with a different difficulty level (performed in random order),
- the control task period that was either before or after the three inverted pendulum periods, but not in the middle (the position was random for each participant).

Each period lasted for five minutes. After each period (including rest), the participant was presented with a self-report questionnaire administered by the experiment supervisor and then the next period began immediately.

One time block was performed with low physical load while the other was performed with high physical load. The order in which the two blocks were presented was randomized. The order of difficulty levels was different for each block, but the control task was always in the same time period for both blocks (either always at the beginning or always at the end). The order of the two blocks and the difficulty levels was randomly chosen before the participant's arrival in the laboratory.

Upon arrival, the participant was told that the purpose of the experiment was to measure the physiological effects of physical load, concentration, stress and habituation. The inverted pendulum task and the experiment procedure were explained, but the participants were deliberately misled about the different difficulty levels present in the task. They were specifically told that there would be no changes in difficulty and that the task would be repeated three times for each level of physical load in order to study the effects of learning and boredom. Then, the inverted pendulum task was presented for the participant to practice at the level of physical load that would be present during the first block. Everyone was required to practice for at least five minutes, and more time was given to anyone who felt that he or she had not yet reached a basic level of proficiency. After practice had been completed, the physiological measurement and gravity compensation equipment was attached

and turned on. The torque output by the gravity compensation system was adjusted manually for each participant. Physiological signals were monitored until the participant became used to the environment and most effects of nervousness or novelty passed. Then, the first block of the experiment was performed. After the first block had been completed, participants were allowed to leave the laboratory for several hours. Upon their return, they were required to practice the inverted pendulum task at the level of physical load that would be present during the second block for at least five minutes. After practice had been completed, the physiological measurement and gravity compensation equipment was reattached and turned on. Physiological signals were monitored until the participant became used to the environment and most effects of nervousness or novelty passed. Then, the second block of the experiment was performed. After the second block had been completed, the participant was disconnected from the equipment and an informal interview was conducted about the entire experience. After this interview, the participant was told about the different difficulty levels present during the task. They were asked not to reveal this information to anyone until the completion of the study. The long pause between the two blocks was due to the length of the experiment; participants during preliminary experiments performed both blocks immediately one after another, but complained about the experiment being too long.

The purpose of the different versions of the inverted pendulum task was to determine how physiological responses are affected by three different things: arousal, valence and physical load. Arousal is defined as a person's level of general mental activity while valence indicates whether the person's emotions are positive or negative (Russell, 1980). By creating six different versions of the same task as well as two different versions of the control task, we hoped to study different combinations of arousal, valence and physical load. Comparing the control task to baseline measurements should show the effects of low and high physical load in the absence of arousal and valence. Comparing the neutral version of the task to the control task should show the added effects of arousal. Comparing the hard version of the task to the neutral version should primarily show the effects of negative valence, though a small increase in arousal was also expected. The negative valence was expected since the task was designed to be too difficult for any participant to handle, thus inducing frustration and stress. The easy version of the task was expected to result in lower arousal as well as slightly different valence than the neutral version. As the participants were explicitly told that there would be no changes in difficulty level, it was expected that some of them would believe that the tasks were easier due to their own improved performance and would therefore feel pride and satisfaction. However, we acknowledged that inducing positive valence would be more difficult than inducing negative valence. While we can reasonably expect most participants to be frustrated by an unreasonably difficult task, they may not derive any satisfaction from apparently performing better. Additionally, it was quite possible that some participants would realize that the difficulty level was being changed during the experiment despite claims to the contrary. This may not decrease the frustration felt during the hard task, but would certainly negate any satisfaction felt during the easy task.

4.4.4 Biomechanical measurements

The position, velocity and acceleration of the HapticMaster as well as the force exerted by participants on the HapticMaster were continuously recorded during the control task and the inverted pendulum task. From this data, six parameters were extracted for each time period: the mean absolute force, the standard deviation of the force signal, the mean frequency of the force signal, the mean frequency of the HapticMaster position signal, the mean frequency of the velocity signal and the mean frequency of the acceleration signal. The mean absolute force was calculated primarily as a method of estimating the physical load exerted by the participant. During preliminary tests, we observed that participants' movement became more sudden, jerky and frantic during the highest difficulty of the inverted pendulum task. The mean frequencies and the standard deviation of the force signal were calculated in an attempt to evaluate this "suddenness" of movement: jerky and irregular movement should be reflected in a higher mean frequency of the movement and force signals.

4.4.5 Participants

Thirty students and staff members of the University of Ljubljana (age range: 19-46 years, mean 26.2, standard deviation 5.8) participated in the study. Twenty-three were male, seven were female. All were healthy, without any major cognitive or physical defects. Four of them already had prior experience with the system. Sixteen had an engineering background but no experience with the setup, and ten had a humanities background and no experience with the setup. Each participant signed an informed consent form.

4.4.6 Data analysis methods

When comparing the inverted pendulum and control tasks to baseline, we used the actual values of the measured physiological parameters for each participant. However, when comparing two different difficulty levels or the pendulum task to the control task, we compared normalized values of physiological parameters for each participant. The normalized value of a physiological parameter was calculated for a particular period by subtracting the baseline value from the value for that period and dividing the obtained difference by the baseline value.

$$x_{relative} = \frac{x_{task} - x_{baseline}}{x_{baseline}} \quad (\text{Eq. 1})$$

The lone exception to this among all physiological parameters was the mean skin conductance level, which is already measured as a deviation from a preset baseline value due to hardware constraints. Thus, for mean SCL the normalized value for a particular period was calculated simply by subtracting the baseline value from the value for that period. Actual values were used for all comparisons of biomechanical parameters since there was no force or movement during the baseline period and thus no way to calculate meaningful normalized values.

For biomechanical measures, actual values of the measured parameters were used to compare different difficulty levels or the pendulum task to the control task. Since there was no interaction between the user and HapticMaster during the baseline

period, all baseline values were zero and it was impossible to calculate normalized values.

When comparing sets of data, statistical significance of differences was calculated using a One-way Repeated Measures ANOVA followed by the Tukey test in post-hoc analysis. If the assumptions for regular ANOVA were not met, ANOVA on Ranks was used instead. The differences were considered statistically significant for $p < 0.05$.

4.4.7 Results – questionnaires

Table 4-3 shows results from the self-assessment manikin for baseline measurements, the control task and the three difficulty levels of the inverted pendulum task. For purposes of analysis, the pictures were assigned numerical values from 1 to 9. On the arousal scale, 1 represented very low arousal while 9 represented high arousal. On the valence scale, 1 represented very negative valence while 9 represented very positive valence. Tables 4-4 and 4-5 show p-values for all pairwise comparisons of self-reported valence.

Self-reported arousal during all three difficulty levels was significantly higher than during the control task or baseline period ($p < 0.001$ for all baseline-task and control-task comparisons). Self-reported arousal was also significantly higher during the control task than during the baseline period in case of high physical load ($p = 0.02$) and approached significance in case of low physical load ($p = 0.08$). The differences in self-reported arousal between the three task difficulty levels were not statistically significant.

Table 4-3 Results of self-report measures for 30 participants, presented as mean +/- standard deviation. Higher values represent higher valence or arousal.

		baseline	control task	difficulty		
				low	medium	high
low physical load	valence	5.5+/-1.1	4.9+/-1.5	5.8+/-1.0	5.4+/-1.2	4.1+/-1.6
	arousal	1.3+/-1.3	1.9+/-1.8	4.6+/-1.8	4.5+/-1.9	4.3+/-1.9
high physical load	valence	5.4+/-1.5	4.5+/-1.9	5.6+/-1.4	5.5+/-1.3	4.5+/-1.8
	arousal	1.3+/-1.7	2.2+/-1.7	4.5+/-1.7	4.7+/-1.7	4.5+/-1.8

Table 4-4 p-values for pairwise comparisons of self-reported valence during low physical load.

	control	low diff.	medium diff.	high diff.
baseline	0.04	0.15	0.89	< 0.001
control		0.001	0.11	0.07
low difficulty			0.11	< 0.001
medium difficulty				< 0.001

Table 4-5 p-values for pairwise comparisons of self-reported valence during high physical load.

	control	low diff.	medium diff.	high diff.
baseline	0.02	0.51	0.81	0.007
control		0.022	0.06	1.00
low difficulty			0.94	0.001
medium difficulty				0.006

4.4.8 Results – biomechanical measurements

For the control task, participants were encouraged to move the HapticMaster at speeds similar to those during the inverted pendulum task. Nonetheless, mean absolute force was significantly higher ($p < 0.001$) during the control task than during any of the three difficulty levels of the inverted pendulum task for both levels of physical load. For low physical load, mean absolute force was 3.1 N +/- 6.9 N (mean +/- standard deviation) for the control task and 1.2 +/- 2.7 N for the inverted pendulum task regardless of difficulty level. The difference was even more drastic for high physical load: 17.1 +/- 6.9 N for the control task and 7.0 +/- 3.4 N for all three difficulty levels of the inverted pendulum task regardless of difficulty level.

Similarly, the standard deviation of the force signal, the mean frequency of the position signal and the mean frequency of the acceleration signal were significantly higher ($p < 0.001$) while the mean frequency of the velocity signal was significantly lower ($p < 0.001$) during the control task than during any difficulty level of the inverted pendulum task.

Introduced primarily as a measure of physical load, mean absolute force was not equal for all three difficulty levels. Figure 4-2 shows mean absolute force as a function of task type for both levels of physical load. It was significantly higher during the high difficulty level than during the low difficulty level for both levels of physical load ($p = 0.001$ for low and $p = 0.025$ for high physical load) and higher during the high difficulty level than during the medium difficulty level for low physical load ($p = 0.043$).

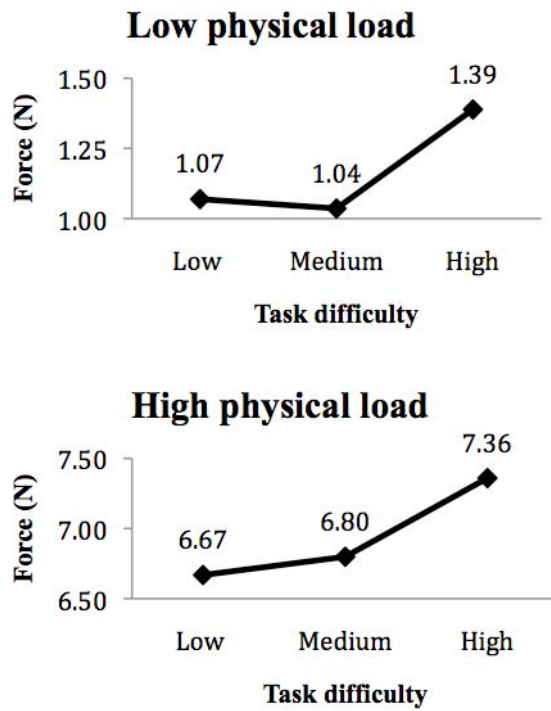


Figure 4-2 Mean absolute force as a function of task difficulty for low and high physical load. Note that values on the vertical axis are not identical on both graphs.

There was no statistically significant difference in any biomechanical parameter between the low and normal difficulty of the inverted pendulum task. Many parameters showed significant differences between the high difficulty level and the other two difficulty levels, but only the mean frequency of the acceleration signal was significantly higher during the high difficulty level for all four pairwise comparisons: high-medium difficulty with low physical load (mean difference 2.4 cm/s², p = 0.015) high-low difficulty with low physical load (mean difference 2.6 cm/s², p = 0.004), high-medium difficulty (mean difference 2.1 cm/s², p = 0.015) with high physical load and high-low difficulty with high physical load (mean difference 1.9 cm/s², p = 0.041).

4.4.9 Results – physiological measurements

Table 4-6 shows the mean normalized values (as defined under the section titled Data analysis methods) of all recorded physiological parameters during the control task and all three difficulty levels of the inverted pendulum task. It also indicates whether the value of each parameter during a task is significantly different from the baseline value.

Table 4-6 Mean values of physiological parameters measured for each task period, with a sample of 30 participants. Results are presented as normalized values (difference from baseline in percentages) except in the case of mean SCL, where results are presented as difference from baseline in microsiemens. ‘Control’ indicates the control task while ‘Low’, ‘Medium’ and ‘High’ indicate different difficulty levels of the inverted pendulum task. Statistically significant differences from baseline value are indicated with bolded values and asterisks: * for $p < 0.05$, ** for $p < 0.01$ and * for $p < 0.001$.**

	Low physical load				High physical load			
	Control	Low	Medium	High	Control	Low	Medium	High
Mean heart rate	-1.6	-1.2	-1.6	-2.4*	10.2***	4.2**	3.9**	4.7***
SDNN	-3.2	-8.3*	-12.3**	-7.4*	-17.1***	-9.3**	-12.3**	-8.8**
RMSSD	-2.7	2.4	2.3	4.0	-23.9***	-6.8	-10.2	-4.1
pNN50	12.2	59.2	29.7	54.1	-26.6**	35.8	27.1	57.1
total HF power	-17.2**	-1.2	-6.4	-6.1	-22.2**	-5.1	-8.5	1.9
total LF power	-8.7*	1.6	-7.4	3.8	3.1	3.8	3.7	16.0
LF/HF ratio	17.0	10.3	14.3	13.6	42.9*	23.5	21.0	29.9
mean SCL (μ S)	0.0	0.6	0.02	0.4	1.0***	0.6**	0.7**	1.0***
SCR frequency	49.2	159.8**	133.7*	175.8*	236.2**	314.0**	340.7**	368.9**
mean respiratory rate	13.7**	16.7***	15.8***	15.2**	17.4***	21.5***	24.0***	22.5***
respiratory rate var.	-14.3**	-	-	-11.1**	-6.0*	-9.2**	-14.4**	4.6
final skin temperature	-0.3	-1.0	-0.1	-1.3*	-0.7	-0.4	-0.1	-1.5*

Percentage-wise, the largest differences between baseline and task were observed for SCR frequency. Normalized values of SCR frequency are shown on Figure 4-3.

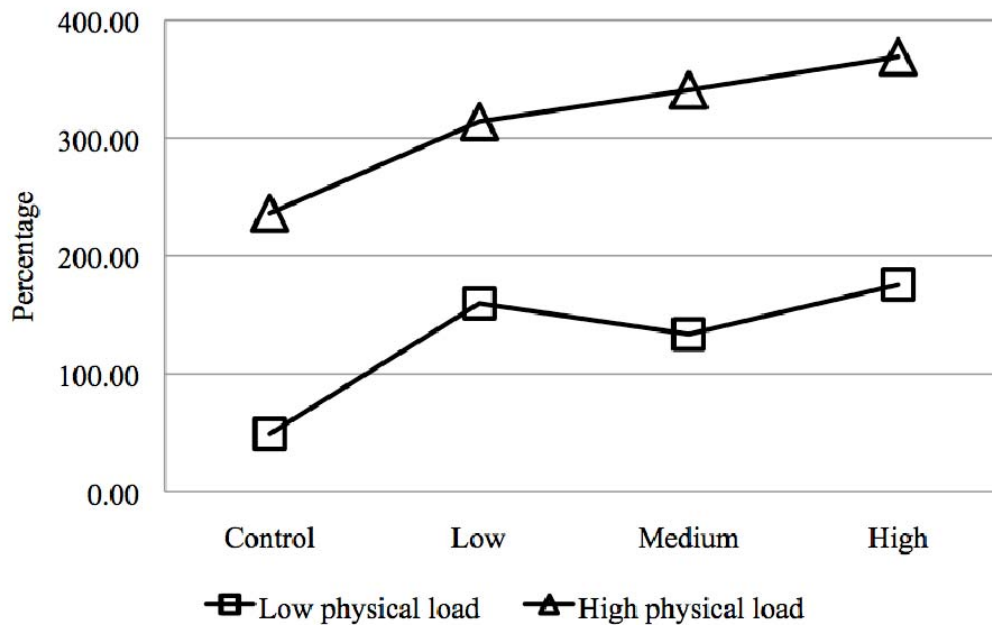


Figure 4-3 Normalized values of skin conductance response frequency as a function of task type. ‘Control’ indicates the control task while ‘Low’, ‘Medium’ and ‘High’ indicate different difficulty levels of the inverted pendulum task.

When studying the effects of arousal and valence, it is not sufficient to only test for significant differences from baseline. To ensure that the differences are not only a result of physical load, we compared normalized values from the inverted pendulum task to the normalized values from the control task. P-values for these comparisons are listed in Table 4-7. In order to gain a better understanding of physiological differences between different task difficulty levels, we also compared normalized values between the three difficulty levels. Most physiological parameters did not show significant differences between difficulty levels, but p-values are listed in Table 8 for all parameters where $p < 0.1$ for at least one comparison.

Table 4-7 Comparison of physiological parameters between the inverted pendulum task and the control task. The normalized value of each physiological parameter during the control task was compared to the normalized value of this parameter during each difficulty level of the inverted pendulum task, and this table shows the p-values for each comparison.

	Low physical load			High physical load		
	Low	Medium	High	Low	Medium	High
Mean heart rate	0.63	0.66	0.88	< 0.001	< 0.001	< 0.001
SDNN	0.31	0.046	0.35	0.16	0.17	0.06
RMSSD	0.41	0.49	0.166	0.002	0.002	< 0.001
pNN50	0.30	0.61	0.024	< 0.001	< 0.001	< 0.001
total HF power	0.027	0.08	0.15	0.16	0.54	< 0.001
total LF power	0.28	0.52	0.11	0.68	0.75	0.21
LF/HF ratio	0.61	0.92	0.77	0.19	0.20	0.75
mean SCL (μ S)	0.020	0.43	0.10	0.39	0.16	0.64
SCR frequency	< 0.001	0.004	0.003	0.09	0.06	0.14
mean respiratory rate	0.015	0.011	0.019	0.015	0.006	0.006
respiratory rate var.	0.875	0.637	0.852	0.005	0.021	0.768
final skin temperature	0.10	0.94	0.20	0.91	0.54	0.32

Table 4-8 Comparison of physiological parameters between different task difficulty levels. Pairwise comparisons were made for normalized values of each physiological parameter during the three difficulty levels, and this table shows the p-values for each comparison. Only physiological parameters where $p < 0.1$ for at least one pairwise comparison are shown.

	Low physical load			High physical load		
	Low-medium	Low-high	Medium-high	Low-medium	Low-high	Medium-high
total HF power	0.90	0.71	0.90	0.47	0.09	0.053
total LF power	0.74	0.44	0.10	0.27	0.016	0.12
respiratory rate var.	0.73	0.048	0.041	0.79	< 0.001	< 0.001
final skin temperature	0.27	0.21	0.007	0.55	0.23	0.011

Respiratory rate variability and final skin temperature were the only two physiological parameters that showed the same differences between tasks both for low and high physical load. To better illustrate differences between tasks, normalized values of respiratory rate variability are shown on Figure 4-4 while normalized values of final skin temperature are shown on Figure 4-5.

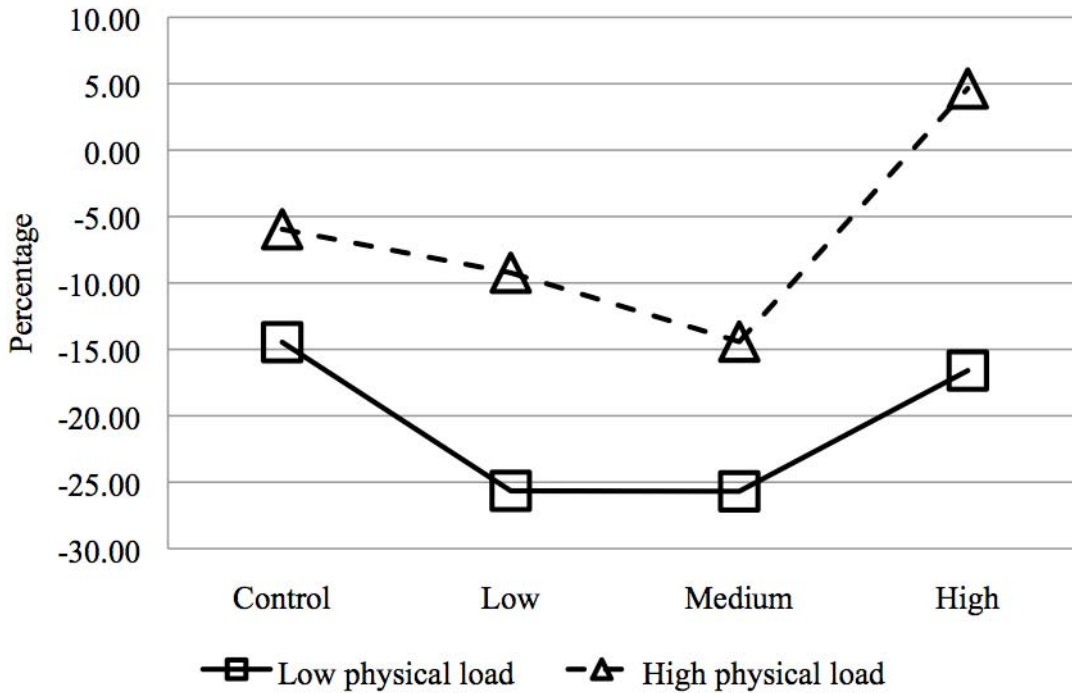


Figure 4-4 Normalized values of respiratory rate variability as a function of task type. 'Control' indicates the control task while 'Low', 'Medium' and 'High' indicate different difficulty levels of the inverted pendulum task.

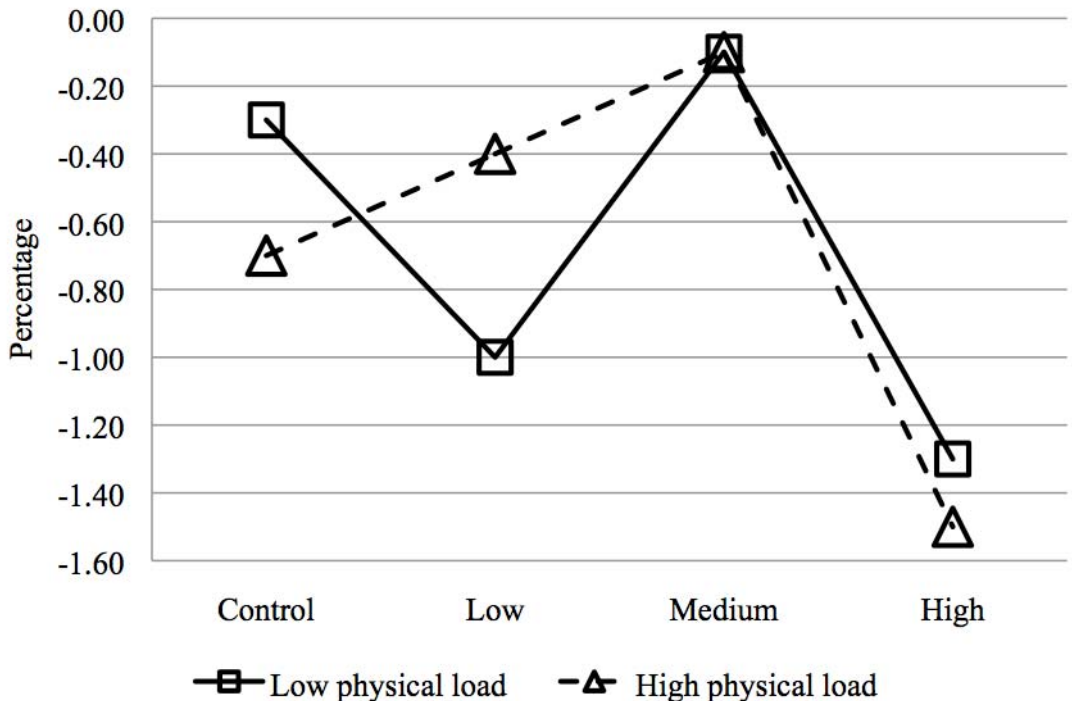


Figure 4-5 Normalized values of final skin temperature as a function of task type. 'Control' indicates the control task while 'Low', 'Medium' and 'High' indicate different difficulty levels of the inverted pendulum task.

4.4.10 Discussion – biomechanical measurements

Looking at results for the mean absolute force, it is immediately apparent that participants were significantly more physically active during the control task than the inverted pendulum task. Additionally, participants were more physically active during the high difficulty level of the inverted pendulum task than during the other two difficulty levels. Thus, we must be cautious when comparing physiological responses to different difficulty levels. The change in a physiological response may not be caused primarily by changes in emotional valence, but instead by the increased physical load associated with increased task difficulty. Physiological measurements from the two different versions of the control task can help us interpret the physiological differences between task difficulty levels.

As predicted, participants' movements during the high difficulty level were more irregular, a fact reflected in significantly higher mean frequency of the HapticMaster acceleration signal. This finding is very specific to the task used in the study. Nonetheless, it may be possible to combine psycho-physiological and biomechanical measures in various applications that require estimation of psychological state and involve physical activity. In tasks where it is possible to identify movement generally associated with changes in arousal or valence, biomechanical measures can provide supplementary information and allow better estimation of psychological state that could be achieved with only psycho-physiological measurements. Additionally, biomechanical measures could help identify whether physiological responses are caused by mental or physical load. For instance, physiological measurements taken during a control task with no mental load and different levels of physical load could be taken before the actual task. Then, during the actual task, biomechanical measures would be used to keep track of current physical load. Using information from the control task, it would be possible to estimate the physiological response caused by physical load and subtract it from the overall physiological response.

4.4.11 Discussion – psycho-physiological measurements

Based on the results of experiment 1, skin conductance and skin temperature were expected to be the most promising parameters. However, both skin conductance level and skin conductance response frequency were shown to be significantly affected by physical load. At the high level of physical load, neither SCL nor SCR frequency were significantly different between the control task and inverted pendulum task. It would seem that skin conductance only serves as a useful indicator of general arousal if not much physical load is required to perform the task. However, it is important to closely examine the levels of physical load that we encountered. As mentioned, the force exerted by the participants during the control task with high physical load was, on average, more than twice the value from the inverted pendulum task. Since the frequency of SCRs increases with physical load, we can assume that it would be much lower in the control task if the mean absolute force exerted were the same as in the inverted pendulum task. In that case, the difference between the control task and the inverted pendulum task would almost certainly be statistically significant; p-values are already lower than 0.15 despite the differences in physical load.

Skin temperature, on the other hand, did appear to be mostly independent of physical load, as it did not significantly change for either version of the control task. It did not decrease significantly for either the low or medium difficulty of the inverted pendulum task, so it cannot be affected purely by arousal. However, it did decrease during the highest difficulty of the inverted pendulum task for both levels of physical load. Since self-reported valence also significantly decreases during the highest difficulty, skin temperature may be an indicator of valence rather than arousal.

Heart rate appeared to be mostly influenced by physical activity, significantly increasing for all task types in the case of high physical load but not in the case of low physical load. Similarly, though changes in heart rate variability were noted, none can be reliably attributed to arousal since they may have been caused by physical load. Respiratory rate, on the other hand, was significantly higher during the inverted pendulum task than during the control task, showing a distinct effect of arousal. Respiratory rate variability appears to decrease as arousal increases, as it was significantly lower during the inverted pendulum task than during the control task. However, it also appears to be influenced by valence – it was significantly higher during the highest difficulty of the inverted pendulum task than during the other two difficulty levels.

4.4.12 Discussion - Inter-subject variability

It is important to note that, though significant psycho-physiological differences were noted between conditions, the variability between subjects was quite large. First of all, subjects can have very different baseline values of psycho-physiological parameters. Skin conductance response (SCR) frequency during rest, for instance, can vary by as much as a factor of five. Using absolute values of psycho-physiological parameters is therefore not feasible. One possible solution is to calculate normalized values by dividing the current value with the baseline value. This results in, for instance, a “20% increase from baseline”.

Nonetheless, considerable inter-subject variability remains even with normalized values. We cannot realistically expect two subjects’ psycho-physiological responses to change in the same way when faced with the same situation. This is partially due to the fact that people cope with situations differently, so the same situation would evoke different emotions in different people. However, we also cannot reliably expect two people to have the same absolute or normalized psycho-physiological response to a situation even if their increase in motivation or engagement is the same. For example, the baseline value of SCR frequency varies among individuals, as has already been mentioned. But there is also a physiological saturation value of SCR frequency, as skin conductance responses cannot happen immediately after another. Thus, a person with a low baseline SCR frequency can exhibit a large normalized increase without reaching physiological saturation (normalized increases of up to 500% have been noted) while a person with a high baseline SCR frequency can only exhibit a small normalized increase (less than 100%) before reaching saturation.

Even within a single person, psycho-physiological responses do not always stay the same. Several studies have noted the process of ‘habituation’, where a person’s

psycho-physiological responses to a certain event lessen as the person is repeatedly exposed to the event.

All the concerns listed above suggest that any decision-making system will have to take the peculiarities of each person into account. At the minimum, a decision-making system will need to know a person's baseline values to be able to function effectively. Additional data such as psycho-physiological responses to a standardized task or results of personality questionnaires may also prove necessary. For the moment, we recommend starting development on a decision-making system that is primarily based on arousal-related parameters. Skin conductance has been repeatedly noted in literature to be influenced only by the sympathetic part of the autonomic nervous system and thus an excellent indicator of arousal. Respiratory rate is also generally agreed to be an indicator of arousal. On the other hand, while respiratory rate variability and skin temperature were indicators of valence in our study, there is not yet widespread agreement about their meaning among psycho-physiologists. Additionally, arousal is most likely easier to control than valence (positive emotions) is. Once such a system has been tested, it can be expanded to control valence as well.

4.4.13 Conclusions

We were able to demonstrate a significant influence of both mental arousal and emotional valence on physiological responses even in the presence of significant physical load. Mean respiratory rate and frequency of skin conductance responses are both indicators of arousal, but also both strongly influenced by physical load. Peripheral skin temperature appears to be independent of the level of physical load and decreases most significantly for situations with negative valence. Respiratory rate variability is affected by both arousal and negative valence. While most of these findings are not new, nearly all previous psycho-physiological studies have been performed in situations with little or no physical load. Our experiment showed that psycho-physiological responses yield useful results even in the presence of physical load and can thus be used in situations such as motor rehabilitation. However, due to the significant effect of physical load, we recommend also using non-physiological sensors (e.g. force or movement sensors) to measure the level of physical load. With this information, it would be possible to estimate the degree to which physiological measurements are affected by physical load and thus make corrections to estimations of psychological state. Furthermore, these sensors may offer additional, task-specific information about a subject's psychological state.

4.5 Experiment 3: Psychophysiology in post-stroke upper limb rehabilitation

4.5.1 Goal

Experiment 2 showed that psycho-physiological responses yield useful results in the presence of physical load. However, this was shown only for healthy subjects. In individuals post stroke, many physiological functions are impaired. For instance, heart rate variability remains depressed several years after a stroke. Therefore, findings from psycho-physiological studies conducted on healthy subjects may not be applicable to stroke patients. The goal of our third experiment was to examine

psycho-physiological responses in both subacute (within a year of stroke) and chronic (several years after stroke) patients.

4.5.2 Tasks

Two different tasks were studied: a purely cognitive task requiring no physical effort and a typical virtual rehabilitation task. For the cognitive task, we used a variant of the classic Stroop colour interference task (Stroop, 1935) that has been previously extensively studied by psychologists. In this task, participants were shown a word on the screen. The word was either “red”, “blue” or “green”. The colour of the word was also either red, blue or green – but the word and the colour of the word did not always match (e. g. **red**, **blue**, **green**). Participants were given a keypad with three grey buttons, with the words “red”, “blue” and “green” written above the buttons in black. They were told to press the button corresponding to the colour of the word (for instance, press blue for **red**). Once a button was pressed, a new word was generated. Occasionally, however, the word was generated in black colour (e. g. **red**). In this case, participants had to push the button corresponding to the word on the screen rather than its colour. The words were randomly generated with the following probabilities:

- word with matching colour (e. g. **red**) – 40%
- word with non-matching colour (e. g. **red**) - 40%
- word in black (e. g. **red**) – 20%

Though simple to implement, this task was nonetheless quite difficult for many participants. It principally served as a test of whether a mentally challenging task would evoke physiological responses in stroke patients. In addition to the Stroop task, a Stroop control task was implemented where a coloured string of X's (e. g. **XXXX**) was shown instead of the word. Participants once again had to push the button corresponding to the colour, but were not distracted by the word. The black colour was not used in the Stroop control task.

The virtual rehabilitation task was designed in our laboratory for reaching and grasping exercise using the HapticMaster. The virtual scenario consisted of a ball rolling down a slope. Participants had to catch the ball as it was rolling, then place it in a basket suspended above the slope. A screenshot of the task display is shown in Figure 4-6.

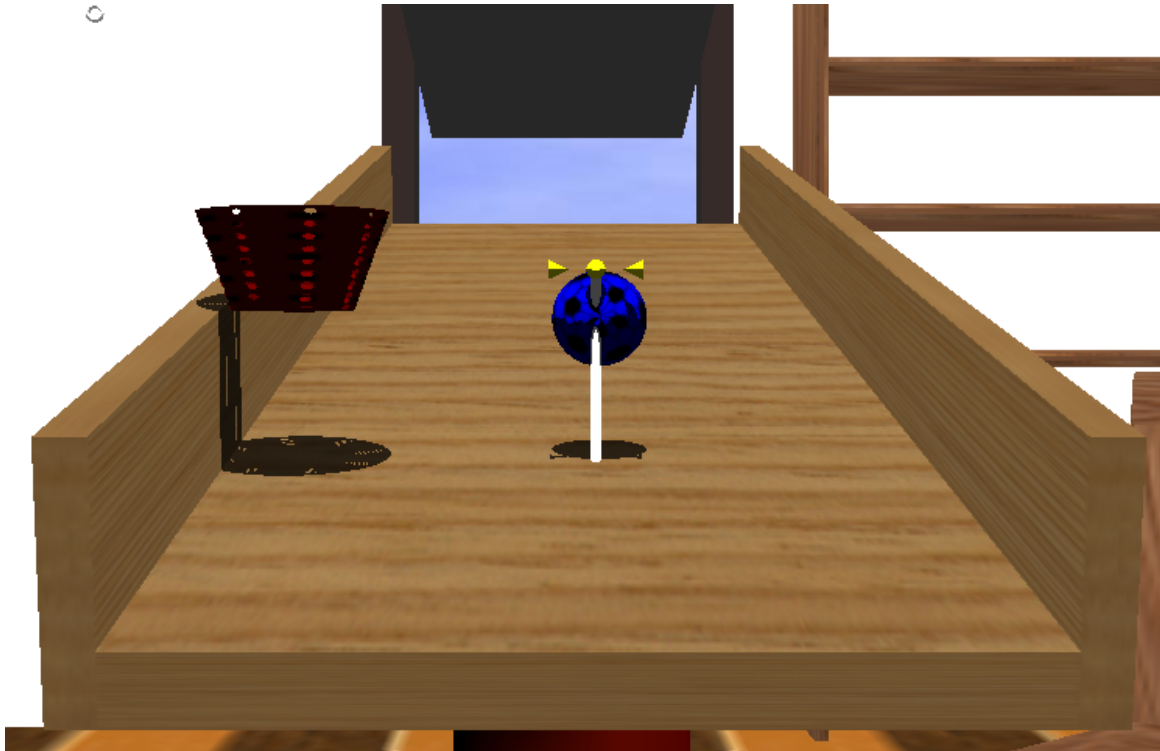


Figure 4-6 The virtual rehabilitation task.

The participant's arm was supported using two cuffs fastened above and below the elbow (as described in deliverable D1.2). These cuffs were connected to a computer-controlled electric motor pulley using Kevlar cables. The pulley applied a constant torque in order to compensate for the gravity acting on the participant's arm. The participant sat approximately 1.5 meters in front of the screen, with the HapticMaster situated between the seat and the screen. The HapticMaster offered various modes of the support. If the participant was unable to perform any or all of the following, the HapticMaster would do it for the participant, guiding his or her arm:

- move left or right in order to catch the ball
- squeeze the grasping device in order to grip the ball
- lift the ball into the basket

This virtual rehabilitation task served as a test of whether a combination of mental and physical effort encountered during a typical rehabilitation scenario would evoke physiological responses in stroke patients. A second, more difficult version of the task was also designed. Meant to be more mentally demanding but equally physically demanding, the difficult version had inverted left-right controls. If the participant moved his or her arm to the left, the virtual hand on the screen moved right (and vice-versa).

In addition to the virtual rehabilitation task, a physical control task was introduced to evaluate the effect of physical load in the absence of mental load. During the physical control task, participants moved the HapticMaster left and right at an even, moderate speed while nothing was shown on the display and all force feedback was disabled.

4.5.3 Experimental procedure

The experiment was conducted in a dedicated, quiet room at the Institute for Rehabilitation of the Republic of Slovenia. Three people were present: the participant, the experiment supervisor and the occupational therapist. Participants were asked by the therapist to fill out the BIS/BAS questionnaire (detailed in section 4.2) prior to arrival. Upon arrival, participants were informed of the purpose and procedure of the experiment, then signed an informed consent form. Then, they were seated in front of the HapticMaster. The afflicted arm was strapped into the arm support cuffs and grasping device. The virtual rehabilitation task was demonstrated, and participants were allowed to practice it for a few minutes. During practice, the HapticMaster was set to perform the parts of the task that participants could not do on their own. After practicing the normal version, a few minutes were also given to practice the more difficult version with the inverted controls.

After the brief practice period, the physiological measurement equipment was attached, turned on and tested. Skin conductance and skin temperature sensors were attached to the unafflicted hand (which was not strapped into the grasping device). Then, participants went through the following procedure:

- rest period (3 minutes)
- physical control task (3 minutes)
- rest period (3 minutes)
- virtual rehabilitation task (3 minutes)
- harder version of the virtual rehabilitation task (3 minutes)

The self-assessment manikin (section 4.2) was presented on the screen after every 3-minute period. Participants were asked to make a selection verbally for both arousal and valence scales.

While the experimenter and occupational therapist did their best to maintain silence in the room during the rest period and control task, the therapist was permitted to verbally guide and encourage the patient during the virtual rehabilitation task. Our reasoning was that verbal communication would increase mental arousal, but not physical load. Since we wished to compare the virtual rehabilitation task to the control task (which had very low mental arousal), increased arousal during the virtual rehabilitation task was considered acceptable. Additionally, verbal guidance was necessary for some patients who had trouble performing the task on their own.

Once the participant had performed both versions of the virtual rehabilitation task, the special keypad used for the Stroop task was strapped to the participant's leg so that it would not fall off during the task. The experimenter explained and demonstrated the Stroop task, and a few minutes were once again given to practice. Participants pushed the buttons on the keypad with the thumb of their unafflicted hand. The skin conductance and skin temperature sensors were not removed, but remained on the other fingers of the unafflicted hand. While this may have affected measurements, it was necessary since many participants were unable to push buttons with the afflicted hand. We felt that moving the sensors to the afflicted arm would have affected measurements even more strongly.

Following the brief practice session, participants rested for three minutes, then performed the Stroop task for three minutes and the Stroop control task for three minutes. Once again, the SAM was presented after every three-minute period. Then, if the participant was tired or needed elsewhere, the experiment was concluded. Otherwise, there was another three-minute rest period followed by both difficulties of the virtual rehabilitation task. This time, the two difficulty levels were presented in random order (the first time round, the easier difficulty level was always first).

4.5.4 Biomechanical measurements

During the data analysis stage, we will evaluate patients' motor abilities using a number of parameters that describe how quickly the patient is able to move in all three dimensions, how well the patient can grasp objects, how 'smooth' the movement trajectories are, whether the patient is better at catching balls on one particular side of the table (as some patients have poorer vision on one side as a result of stroke) etc. Knowledge gained from previous studies of reaching and grasping exercise will be used as a basis.

4.5.5 Participants

While the experiment is still in progress, we are planning to include 20 subacute stroke patients and a group of 20 age- and gender-matched healthy control subjects. Chronic patients will be included subject to availability. Each participant will be asked to perform the experiment twice, with approximately two weeks passing between the first and second visit.

As of June 16th, 2009, fourteen subacute patients and four chronic patients have participated in the experiment. Results from ten of the subacute patients have been analyzed.

4.5.6 Results

Although the measurements have not yet been completed, psychophysiological results from the first ten patients are promising. Distinct physiological responses have been noted both in response to distinct events (e.g. surprise) and as a result of general activity. The Stroop task in particular triggers strong psychophysiological responses that cannot be attributed to physical activity. Figures 4-7 to 4-10 present a brief overview of the psychophysiological results obtained thus far. Differences between baseline and task are shown for three periods: the movement task, the very first virtual rehabilitation task (easier version, performed after the movement task) and the Stroop task. On each figure, the ends of the boxes indicate the 25th and 75th percentiles, with a line at the median and error bars defining the 10th and 90th percentiles. The dots outside the error bars indicate outliers. Though the spread of the results is quite large, Wilcoxon signed-rank tests have already shown significant differences in a number of psychophysiological parameters between baseline and task for the virtual rehabilitation task and the Stroop task.

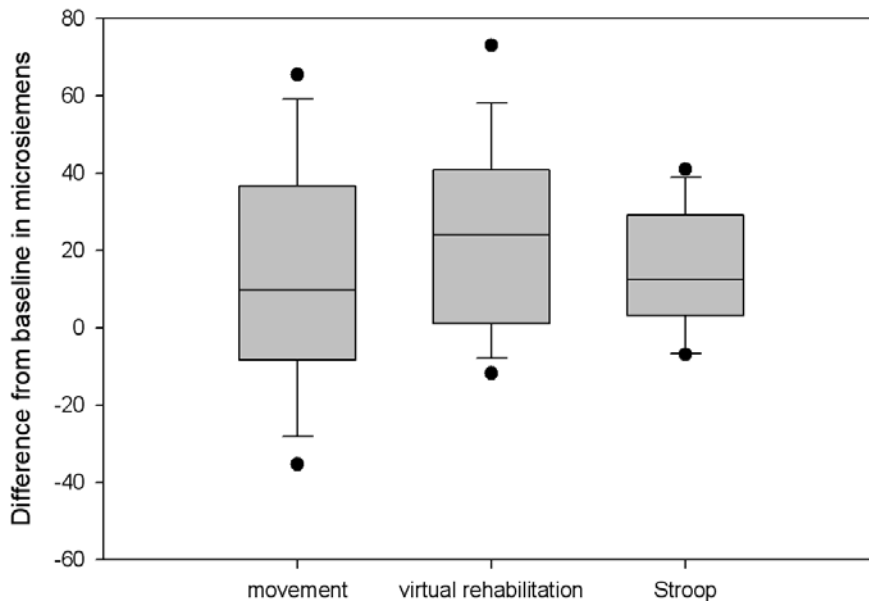


Figure 4-7: Preliminary skin conductance level results for 10 stroke patients.

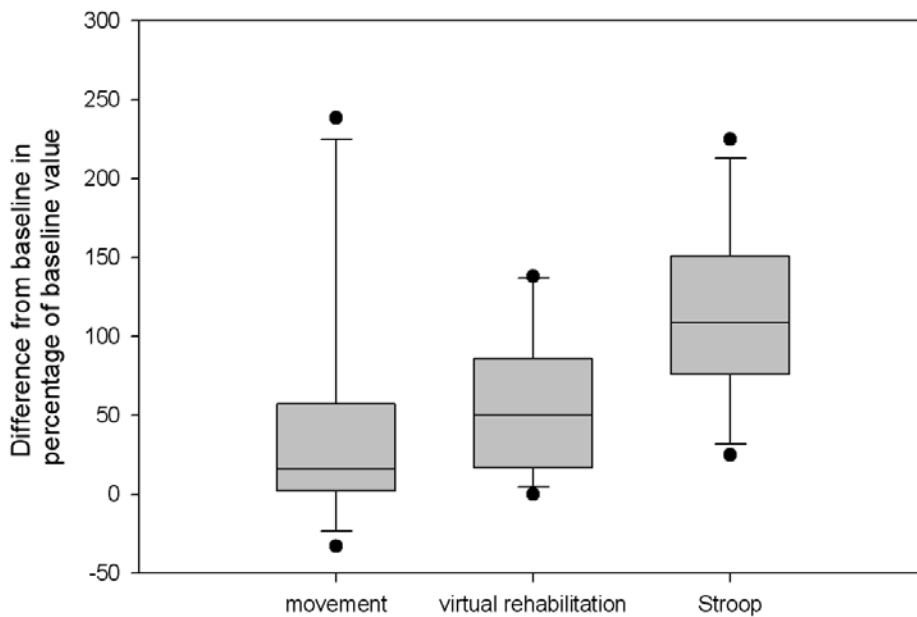


Figure 4-8: Preliminary skin conductance response frequency results for 10 stroke patients. Omitted: one outlier at 400% for movement and one outlier at 1200% for virtual rehabilitation.

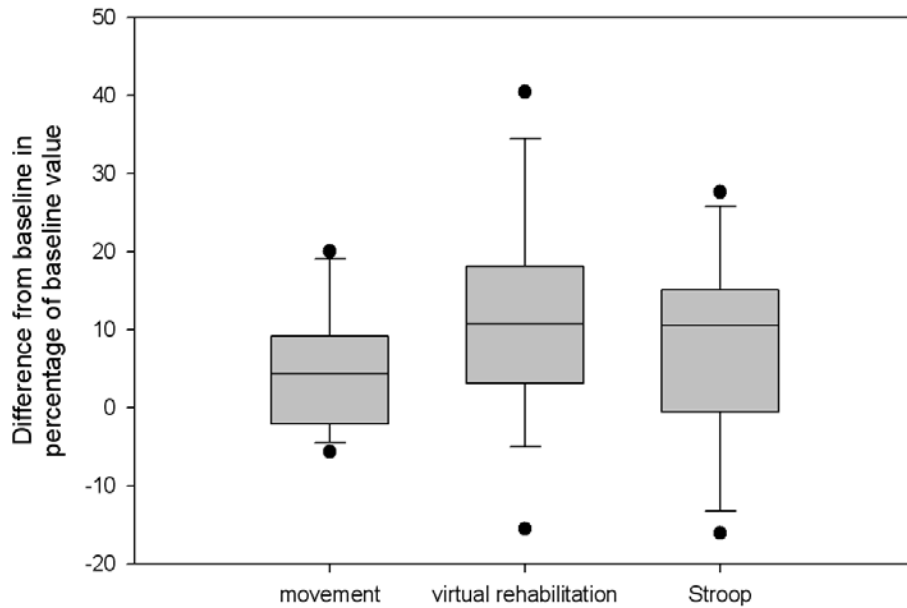


Figure 4-9: Preliminary respiratory rate results for 10 stroke patients.

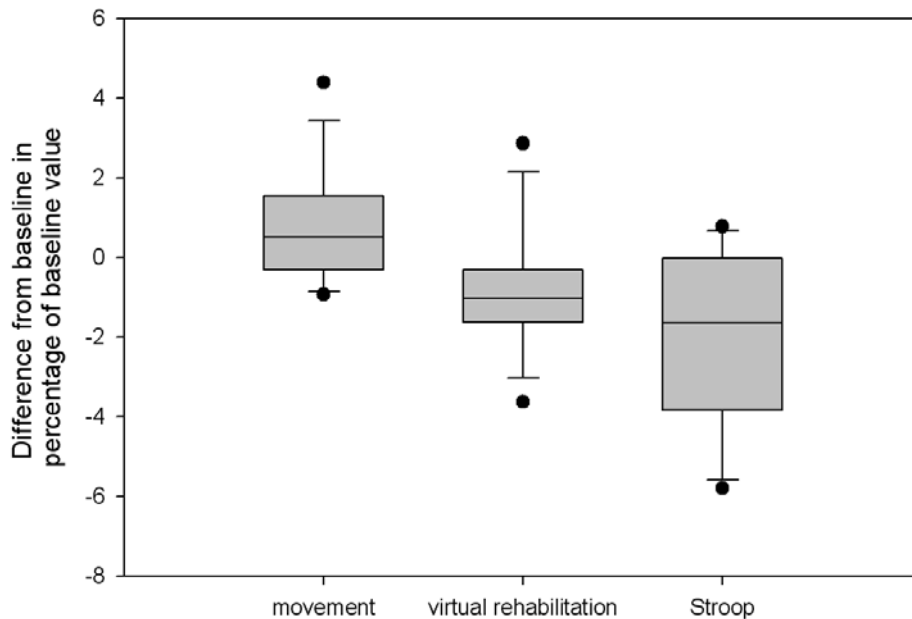


Figure 4-10: Preliminary skin temperature results for 10 stroke patients.

However, it will be necessary to measure and analyze the entire 20-patient group before we can reliably determine whether psychophysiological responses during the virtual rehabilitation task are caused mainly by the effort of moving the arm or also triggered by visual and acoustic stimuli from the virtual environment. While the previous experiment showed that arousal from the virtual environment substantially contributes to psychophysiological responses of healthy subjects even in the presence of physical activity, stroke patients may find physical activity far more demanding and arousing. Additionally, the entire 20-patient group will need to be

measured before we will be able to determine whether psychophysiological responses depend on the type and location of stroke, the amount of time that has passed since the stroke, the degree of motor recovery etc.

One important finding has been that use of the self-assessment manikin with stroke patients in a clinical setting can be very problematic. A number of patients always chose the same valence and arousal values throughout the experiment (e.g. lowest possible arousal during rest, during the Stroop task and during the virtual rehabilitation task). This does not appear to be dependent on a single factor. It has been noted both in patients with lesions in the left hemisphere and patients in the right hemisphere, in patients who can otherwise talk normally and express emotions such as annoyance and amusement (some actually verbally admitted that changes in arousal or valence had occurred when prompted verbally by the experiment supervisor, but refused to change their answer to the SAM) etc. In some patients, claiming high valence and low arousal may have simply been a way of trying to please the experimenter and therapist by not appearing overworked or disinterested. This weakness of self-evaluation questionnaires strengthens the need for an objective measure of patient state that can be provided by psychophysiology.

5 Conclusion and Outlook

The final goal of the MIMICS project is to enhance the clinical outcome of robot-assisted rehabilitation by increasing patient motivation and participation through the use of novel technology. To assess both the success of the project as a whole and to adapt to the individual patients needs, tractable instruments of measuring the participation and motivation have to be developed.

The project so far has succeeded in providing sensors and systems for measuring both psycho-physiological and biomechanical data and algorithms to assess and differentiate subjects' physical and mental load during various tasks both in the HapticMaster and the Lokomat.

First analyses correlating these results to emotional valence and activation / motivation have been conducted, with slight difficulties regarding patients' ability to use the measurement instruments (specifically the self-assessment manikin). Here, a workable solution still has to be found. One possible alternative instrument to assess task-specific motivation could be the Intrinsic Motivation Inventory (IMI, Ryan et al 1982, 1985), well validated for several medical and psychiatric patient collectives and in sports. Experiments using this scale and a similarly well-validated German questionnaire used to assess the "baseline" motivation of patients in a rehabilitation setting, (PAREMO, Hafen et al., 2001 and PAREMO-20, Kriz et al, 2006) to assess the motivation of patients treated on the MIMICS vs. the conventional Lokomat will start in NKBA in July 2009. These scales aim to take into account the hidden agendas and observer effects that could be part of the problems encountered with SAM and BIS / BAS.

We feel that significant progress has been made in all subtasks targeted by this Deliverable, transfer to the clinical partners for experiments with patients has been initiated, and therefore Deliverable 4.2 can be considered fully achieved as envisioned by the quality assurance plan.

6 References

Carver, C. S. & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, 67, 319-333.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119-137). Norwood, NJ: Ablex.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, pp 1161–1178.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, pp 643-662.

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17, pp 354-381.

Witmer, B. & Singer, M. (1998). Measuring Presence in Virtual Environments: a Presence Questionnaire. *Presence*, 7, pp 225-241.

Ryan, R.M., (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43, 450-461.

Ryan, R.M. (1985). Intrinsic motivation and the effects of self-consciousness, self-awareness and ego-involvement: An investigation of internally-controlling styles. *Journal of Personality*, 53, 435-449.

Hafen, K., Jastrebow, J., Nübling, R., Bengel, J. (2001). Entwicklung eines Patientenfragebogens zur Erfassung der Reha-Motivation (PAREMO). *Rehabilitation*, 40, 3-11.

D Kriz, R Nübling, JE Herwig, S Fuchs, J Bengel (2006). Normierung des PAREMO-20 - Patientenfragebogen zur Erfassung der Rehabilitationsmotivation. *DRV-Schriften*.