Robotica: page 1 of 8. © Cambridge University Press 2010
doi:10.1017/S0263574710000184

Psychophysiological responses to different levels of cognitive and physical workload in haptic interaction

Domen Novak*, Matjaž Mihelj and Marko Munih

University of Ljubljana, Faculty of Electrical Engineering, Trzaska cesta 25, 1000 Ljubljana, Slovenia

(Received in Final Form: April 16, 2010)

SUMMARY
Psychophysiological measurements, which serve as objective indicators of psychological state, have recently been introduced into human–robot interaction. However, their usefulness in haptic interaction is uncertain, since they are influenced by physical workload. This study analyses psychophysiological responses to a haptic task with three different difficulty levels and two different levels of physical load. Four physiological responses were recorded: heart rate, skin conductance, respiratory rate and skin temperature. Results show that mean respiratory rate, respiratory rate variability and skin temperature show significant differences between difficulty levels regardless of physical load and can be used to estimate cognitive workload in haptic interaction.

KEYWORDS: Haptic interfaces; Psychophysiology; Virtual reality; Human factors; Multimodal interfaces.

1. Introduction

Changes in a person's psychological state are reflected by the person's physiological state. For instance, stressful situations cause increased sweating and changes in heart rate. These responses are generally modulated by the sympathetic branch of the autonomic nervous system and lead to the body experiencing what is commonly termed stress. When the causative conditions change and the body can recover, the parasympathetic branch of the autonomic nervous system reduces the body's stress level in an attempt to revert the body back to its normal state.[1] These physiological responses to psychological states can be measured using the so-called psychophysiological measurements. Such measurements can be taken without the subject's active cooperation, providing a convenient, objective and unobtrusive method of estimating a person's psychological state. Because of these advantages, they have been used in a variety of situations, including virtual reality[2,3] and human–robot interaction.[4–6]

From a psychophysiological perspective, most situations in human–robot interaction do not present any major difficulties. However, haptic interaction can present a special challenge. When using a haptic robot, users may need to exert very large forces and torques. Most psychophysiological studies, on the other hand, have focused on tasks and situations that require little physical activity. This has been primarily because it is difficult to separate the physiological effects of physical and cognitive (mental) workload, a problem that has been noted in recent applied studies.[7] Although a number of studies have examined psychophysiological responses to a combination of physical and cognitive workload, they have mainly focused on the effects of a mentally demanding task superimposed onto a physically demanding task (e.g. performing mental arithmetic while riding a bicycle).[8,9] Subjects in these studies were thus performing several unrelated tasks at once. During interaction with haptic robots, however, a single task frequently contains elements of both physical and cognitive workload. The interplay between cognitive and physical workload found in haptic human–robot interaction may result in different psychophysiological responses.

The question we wished to answer was simple: in haptic human–robot interaction, is it possible to use psychophysiological responses to differentiate between different levels of cognitive workload at different levels of physical workload? Our focus was primarily on measurements of the autonomic nervous system, for which the required equipment is relatively inexpensive and easy to apply to the subject. Four psychophysiological responses were examined: heart rate, skin conductance, respiratory rate and peripheral skin temperature. All four provide information about cognitive workload in situations with no physical load, and several are significantly affected by cognitive workload in multi-task situations. Our hypothesis was that, in haptic human–robot interaction, at least some psychophysiological responses would be able to differentiate between different levels of cognitive workload regardless of the level of physical load. This would make them suitable for use in applications where a haptic robot interacting with a human operator could change its level of autonomy based on the operator's psychophysiological state.

2. Materials and Methods

2.1. Hardware and software

The hardware configuration comprised three major parts: the display system, the haptic robot and the signal recording system (Fig. 1). A 2 × 1.5-meter screen with back projection was used to display visual data. The HapticMaster, a high-performance force-controlled robot developed by Moog FCS, was used as the haptic robot. This robot has three degrees of freedom. The first joint allows vertical translation, the second allows rotation around a vertical axis and the third allows

* Corresponding author. E-mail: domen.novak@robo.fe.uni-lj.si

Fig. 1. The hardware configuration and virtual scenario. The subject is seated in front of a screen (centre) and manipulates the HapticMaster (1) with his/her dominant arm. The screen displays the inverted pendulum balancing task with the virtual cart (2) and pole (3). The circles on the left and right show the sensors used. The force and position sensors are integrated into the HapticMaster while the physiological sensors are attached to the subject.

horizontal translation. The robot's end-effector also contains sensors for measurement of forces in three dimensions. Further information is available in an extensive paper by van der Linde & Lammertse.[10] The robot's end-effector was held by the subject with his or her hand, allowing movement of the robot in three dimensions. For purposes of our task, the robot was limited to left–right movement. All force and movement data were recorded at a sampling frequency of 2.5 kHz. The subject's arm was supported using two cuffs fastened above and below the elbow. These cuffs were connected to electric motor pulleys using Kevlar cables. The pulleys applied a constant torque in order to compensate for the gravity acting on the subject's arm. The subject sat approximately 1.5 meters in front of the screen, with the HapticMaster situated between the seat and the screen.

The electrocardiogram 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 phalanxes of the second and third fingers of the non-dominant hand. The sensor generated a constant voltage between the two electrodes 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.*[11] Respiratory rate was obtained using a thermistor-based SleepSense Flow sensor. Placed beneath the nose, this sensor measured 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. 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.

Both visualization and signal processing were implemented in Matlab/Simulink. Physiological signals were imported directly into Simulink using drivers and Simulink blocks provided by the manufacturer of the signal amplifier. They were recorded raw, then filtered and analysed offline. xPC Target 3.3 was used to control the HapticMaster.

2.2. Subjects

Thirty students and staff members from various departments 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. Each subject signed an informed consent form.

2.3. Task

Subjects were presented with a virtual version of the classic inverted pendulum problem (visible on the screen in Fig. 1). 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; left alone, the pole will fall to the ground. However, if the cart is moved left or right, it will act upon the pole and either accelerate its fall or balance it. This system is referred to as the inverted pendulum and is a classic problem in control theory. Subjects were presented with a simulated cart and pole on a screen and the cart was moved using the HapticMaster, with the goal of keeping the pole from falling. The cart moved in the same direction and with the same velocity as the end-effector of the HapticMaster. If the subjects failed to balance the pole and it fell to a horizontal position, it was immediately reset to a nearly vertical position. Force feedback was also implemented with the HapticMaster, allowing the subjects to feel the reaction forces resulting from the movement of the cart.

Different levels of cognitive workload were achieved in our task using three different task difficulty levels: underchallenging, challenging and overchallenging. In the underchallenging version, the pendulum was not affected by gravity and thus never fell. The subject was simply asked to move the cart left and right at a moderate speed. In the challenging version, the model dynamics were balanced in such a way as to make the balancing of the pendulum moderately challenging. In the overchallenging version, a half-second delay was introduced between the time the cart was moved and the time the cart's movement affected the pole. Additionally, the gravity acting on the pole was stronger and the pole was less responsive to the movement of the cart. This made the task extremely difficult to perform successfully.

All the three difficulty levels were implemented in low and high physical load versions. The versions were identical except for one factor: in the high physical load versions, more physical force was required to move the HapticMaster. The force applied by the subject was divided by five, forcing the subject to apply five times the force that had been applied in the low physical load versions. This gave us a total of six task conditions: underchallenging with low physical load, challenging with low physical load, overchallenging with low physical load, underchallenging with high physical load, challenging with high physical load and overchallenging with high physical load.

2.4. *Experiment protocol*

The experiment was conducted in a quiet area of the laboratory where external stimuli did not disturb the subjects. The temperature and humidity in the laboratory were kept constant. There was never more than one subject and one experiment supervisor inside the laboratory at any time. Each subject performed the experiment in two separate time blocks. Each block consisted of an initial rest period (which served as the baseline) followed by the three different difficulty levels performed in random order. Each condition lasted for five minutes. After each condition, the subject was presented with a self-report questionnaire administered by the experiment supervisor and then the next condition 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 as well as the order of difficulty levels within each block was randomly chosen before each subject's arrival in the laboratory.

Upon arrival, the task and the experiment procedure were explained to the subject. Then, the challenging difficulty level was presented for the subject to practice using the HapticMaster 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. This practice period was presented in order to reduce the effect of novelty: psychophysiological responses are generally strongest during the first exposure to a new stimulus, then decrease as a result of habituation.[1] For this reason, psychophysiological studies frequently perform a practice session before the actual experiment in order to reduce the effects of novelty during the experiment session.[2] Nonetheless, it should be noted that this is not the only possible approach. Some researchers prefer to use a different task for the practice period so that novelty is preserved and the results of the actual experiment are not affected by learning or habituation.[3] Additionally, we wish to point out that, since the practice was performed before the sensors were attached, the subjects' mood during the practice period may have been different than during the experiment itself (when the same task was performed with physiological sensors attached). While we do not believe that this affected the validity of the results, it may be preferable to attach the physiological sensors to the subject before the practice.

After practice, the physiological sensors were attached and turned on. Then, the first block of the experiment was performed. After the first block of experiment had been completed, a brief informal interview was conducted with the subject. He or she was allowed to rest briefly, if desired. Then, he or she was required to practice the task for at least five minutes at the level of physical load that would be present during the second block. After the practice, the second block of the experiment was performed. After the second block had been completed, the subject was disconnected from the equipment and an informal interview was conducted about the entire experience.

2.5. *Performance and self-report measures*

Subjects who are better at the task are able to balance the pendulum and keep it from falling. Thus, in the challenging and overchallenging conditions, the number of times that the pendulum fell (and was reset) was counted and used as a quantitative measure of the subjects' task performance. This was not done for the underchallenging condition where the pendulum remained upright no matter what the subject did.

Subjects evaluated their own psychological state using nine-point arousal and valence scales from the self-assessment manikin.[12] These scales allow the subject to rate their level of emotional valence and arousal graphically by choosing the picture they feel best represents their current mood. Valence (sometimes also called pleasure) is defined as positive versus negative affective states (e.g., excitement, relaxation and tranquility versus cruelty, humiliation, disinterest and boredom) while arousal is defined in terms of mental alertness and physical activity (e.g., sleep, inactivity, boredom and relaxation at the lower end versus wakefulness, bodily tension, strenuous exercise and concentration at the higher end).[13] The questionnaire was given to each subject and the meaning of each scale was explained before the experiment began. The questionnaire was then presented after each baseline and task period. Although the focus of our experiment was on psychophysiological measurements, the questionnaire was used to check whether our subjects really found the three task difficulty levels significantly different from each other. A two-scale questionnaire was used so that the nature of the differences (arousal or valence) between the difficulty levels could be determined.

2.6. *Force measures and physiological measures*

The position of the HapticMaster and the force exerted by subjects on the HapticMaster were continuously recorded. Additionally, each subject's physiological signals were recorded. After the experiment, the signals were band-pass-filtered offline and the relevant parameters were extracted for each time period. From the position and force data, mean absolute force was calculated as a method of estimating the physical load exerted by the subject. From the four physiological signals, a number of different psychophysiological parameters were extracted for each time period. In the rest of the text, all of the psychophysiological parameters analysed in our study are printed in italics.

Analysis of the ECG began by calculating the times between two normal heartbeats (NN intervals). *Mean heart rate* was calculated from the NN intervals, and several standardized measures[14] were used to estimate the 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, NN intervals were converted into an instantaneous time series with a constant sampling frequency using cubic spline interpolation. 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 (referred to as *the LF/HF ratio*). Previous studies have shown that heart rate increases with cognitive workload[7,15] and that it is higher in multi-task situations with both physical and cognitive load than in single-task situations with equivalent physical load but no cognitive load.[6] Decreases in HRV have also been linked to increases in cognitive workload.[15,16]

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 discrete environmental event. Its mean value (*mean SCL*) was calculated. 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. An increase in skin conductance was classified as a SCR if its amplitude exceeded 0.05 μS and the peak occurred less than five seconds after the beginning of the increase. *SCR frequency* was calculated by dividing the total number of detected SCRs with the length of the period (which was always five minutes). Both SCL[17,18] and SCR frequency[19,20] have been found to increase with general psychological arousal and cognitive workload.

Mean respiratory rate was calculated in breaths per minute. 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. Respiratory rate has been found to increase with general arousal and cognitive workload.[15,21] In multi-task situations with both physical and cognitive load, respiratory rate has been shown to be higher than in single-task situations with equivalent physical load but no cognitive load.[9] Respiratory variability in general has been shown to decrease with increased cognitive workload.[22]

Final skin temperature was calculated by averaging temperature during the last five seconds of each time period. Skin temperature has been shown to decrease as a result of cognitive load[23] as well as a result of tension or anxiety.[24]

Although changes in psychological state naturally do occur in the course of each five-minute time period, each parameter was calculated over the entire period. This is because some parameters (e.g. HRV, SCR frequency) can only be calculated over a period of several minutes while others (e.g. skin temperature) change relatively slowly in response to psychological changes.

2.7. *Data analysis methods*

When comparing a task condition to baseline (rest period), we used the actual values of the measured physiological parameters for each subject. However, when comparing different task conditions, we compared relative values of physiological parameters for each subject. The relative value of a physiological parameter was calculated for a particular time period by subtracting the baseline value from the value for that period and dividing the obtained difference by the baseline value:

$$x_{relative} = \frac{x_{task} - x_{baseline}}{x_{baseline}} \tag{1}$$

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

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. Differences were considered statistically significant for $p < 0.05$.

3. Results

3.1. *Performance and self-report measures*

For low physical load, the pendulum was reset 3.2 $\pm$ 1.3 times per minute during the challenging condition (mean $\pm$ standard deviation across all 30 subjects) and 5.6 $\pm$ 1.0 times per minute in the overchallenging condition. For high physical load, the pendulum was reset 2.8 $\pm$ 0.9 times per minute in the challenging condition and 5.4 $\pm$ 1.2 times per minute in the overchallenging condition. Differences between challenging and overchallenging conditions were significant ($p < 0.001$) for both levels of physical load. Differences between low and high physical load were not

Table I. Results of self-report measures, presented as mean +/− standard deviation. High values represent positive valence or high arousal. Underch. = underchallenging, ch. = challenging, overch. = overchallenging.

		Baseline	Underch.	Ch.	Overch.
Low physical load	Valence	5.5+/−1.1	4.9+/−1.5	5.4+/−1.2	4.1+/−1.6
	Arousal	1.3+/−1.3	1.9+/−1.8	4.5+/−1.9	4.3+/−1.9
High physical load	Valence	5.4+/−1.5	4.5+/−1.9	5.5+/−1.3	4.5+/−1.8
	Arousal	1.3+/−1.7	2.2+/−1.7	4.7+/−1.7	4.5+/−1.8

significant for either difficulty level. The pendulum did not fall during the two underchallenging conditions.

Table I shows results from the self-assessment manikin for baseline and task conditions. 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.

Self-reported arousal in the challenging and overchallenging conditions was higher than during the baseline period or the underchallenging conditions for both levels of physical load ($p < 0.001$ for all comparisons). Arousal in the underchallenging condition was only significantly higher than baseline in case of high physical load, though the difference did approach significance in case of low physical load ($p = 0.08$).

For both low and high physical load, self-reported valence during the underchallenging and overchallenging conditions was lower than baseline ($p < 0.05$ for all comparisons). For both low and high physical load, it was lower during the overchallenging than during the challenging condition ($p < 0.01$ for both comparisons).

3.2. *Force measurements*

Introduced primarily as a measure of physical load, mean absolute force was not equal for all the three difficulty levels. For low physical load, mean absolute force was 3.1 ± 6.9 N in underchallenging, 1.0 ± 2.7 N in challenging and 1.4 ± 3.6 N in overchallenging condition. For high physical load, it was 17.1 ± 6.9 in underchallenging, 6.8 ± 3.5 N in challenging and 7.4 ± 2.9 N in overchallenging condition. For both levels of physical load, it was significantly higher in the underchallenging condition than in the other two conditions ($p < 0.001$ for all comparisons). For high physical load, it was also higher in the overchallenging than the challenging condition ($p = 0.025$).

3.3. *Physiological measures*

Table II shows the mean relative values (as defined under the section titled data analysis methods) of all recorded physiological parameters in all task conditions. Bold values and asterisks indicate whether the value of each parameter is significantly different from the baseline value. Percentage-wise, the largest differences between baseline and task were observed for *SCR frequency*.

The physiological effects of physical load were evaluated by comparing relative values between the low and high physical load conditions for each difficulty level (the underchallenging task with low physical load was compared to the underchallenging task with high physical load and so on). However, most physiological parameters did not show a significant difference between low and high physical load. The three exceptions were *mean heart rate*, *mean SCL* and *SCR frequency*. For each of these three parameters, $p < 0.01$ for all the three difficulty levels.

p-values for comparisons of relative values between difficulty levels are listed in Table III. *Respiratory rate variability* and *final skin temperature* were the only two

Table II. Mean values of recorded physiological parameters in all task conditions. Results are presented as relative values (difference from baseline in percentages) except in the case of *mean SCL*, where results are presented as difference from baseline in microsiemens. Statistically significant differences from baseline value are indicated with bolded values and asterisks: * for $p < 0.05$, ** for $p < 0.01$ and *** for $p < 0.001$. Underch. = underchallenging, ch. = challenging, overch. = overchallenging.

	Low physical load			High physical load		
	Underch.	Ch.	Overch.	Underch.	Ch.	Overch.
Mean heart rate	−1.6	−1.6	**−2.4***	**10.2*****	**3.9****	**4.7*****
SDNN	−3.2	**−12.3****	**−7.4***	**−17.1*****	**−12.3****	**−8.8****
RMSSD	−2.7	2.3	4.0	**−23.9*****	−10.2	−4.1
pNN50	12.2	29.7	54.1	**−26.6****	27.1	57.1
Total HF power	**−17.2****	−6.4	−6.1	**−22.2****	−8.5	1.9
Total LF power	**−8.7***	−7.4	3.8	3.1	3.7	16.0
LF/HF ratio	17.0	14.3	13.6	**42.9***	21.0	29.9
Mean SCL (μS)	0.0	0.02	0.4	**1.0*****	**0.7****	**1.0*****
SCR frequency	49.2	**133.7***	**175.8****	**236.2*****	**340.7*****	**368.9*****
Mean respiratory rate	**13.7*****	**15.8*****	**15.2*****	**17.4*****	**24.0*****	**22.5*****
Respiratory rate var.	**−14.3****	**−28.6*****	**−11.1****	**−6.0***	**−14.4****	4.6
Final skin temperature	−0.3	−0.1	**−1.3***	−0.7	−0.1	**−1.5***

Table III. Comparison of physiological parameters between different difficulty levels. Pairwise comparisons were made for relative values of each physiological parameter during the three difficulty levels, and this table shows the *p*-values for each comparison. Underch. = underchallenging, ch. = challenging, overch. = overchallenging.

	Low physical load			High physical load		
	Underch.–Ch.	Underch.–Overch.	Ch.–Overch.	Underch.–Ch.	Underch.–Overch.	Ch.–Overch.
Mean heart rate	0.66	0.88	0.32	< **0.001**	< **0.001**	0.65
SDNN	**0.046**	0.347	0.14	0.167	0.057	0.23
RMSSD	0.49	0.166	0.59	**0.002**	< **0.001**	0.09
pNN50	0.61	**0.024**	0.40	< **0.001**	< **0.001**	0.17
LF/HF ratio	0.92	0.77	0.81	0.20	0.75	0.88
Total HF power	0.08	0.15	0.90	0.54	< **0.001**	0.053
Total LF power	0.52	0.11	0.10	0.75	0.21	0.12
Mean SCL (μS)	0.43	0.10	0.32	0.16	0.64	0.19
SCR frequency	**0.004**	**0.003**	0.053	0.06	0.14	0.49
Mean respiratory rate	**0.011**	**0.019**	0.57	**0.006**	**0.006**	0.57
Respiratory rate var.	0.637	0.852	**0.041**	**0.021**	0.768	< **0.001**
Final skin temperature	0.94	0.20	**0.007**	0.54	0.32	**0.011**

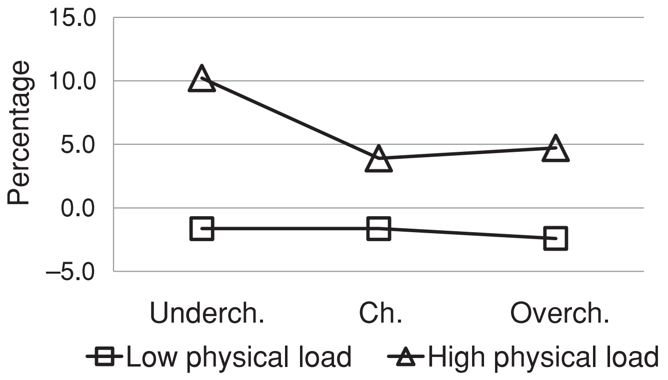

Fig. 2. Mean relative values of *mean heart rate* as a function of task difficulty. Underch. = underchallenging, ch. = challenging, overch. = overchallenging.

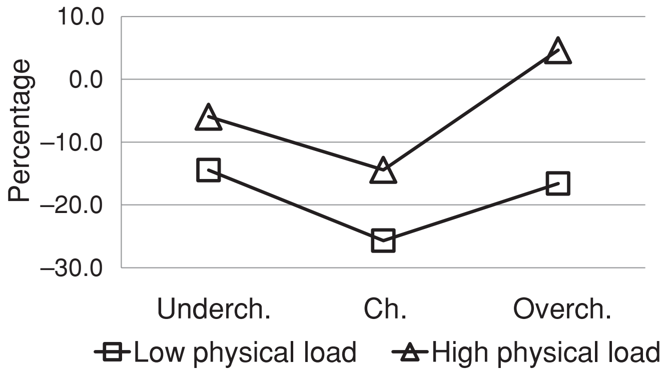

Fig. 4. Mean relative values of *respiratory rate variability* as a function of task difficulty. Underch. = underchallenging, ch. = challenging, overch. = overchallenging.

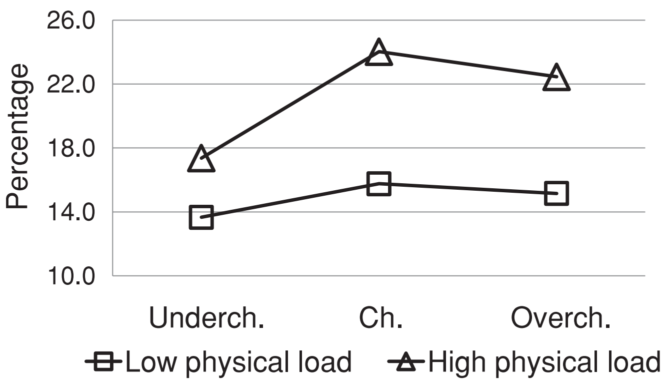

Fig. 3. Mean relative values of *mean respiratory rate* as a function of task difficulty. Underch. = underchallenging, ch. = challenging, overch. = overchallenging.

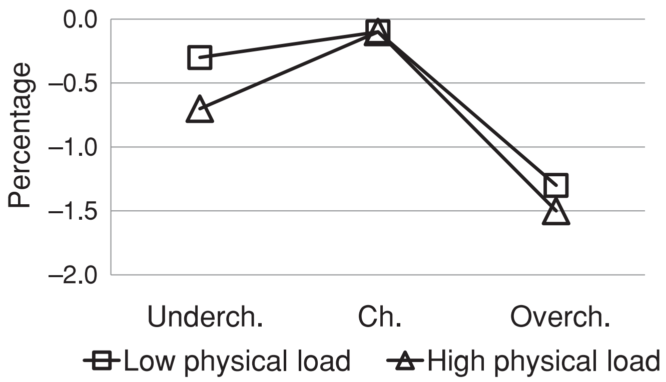

Fig. 5. Mean relative values of *final skin temperature* as a function of task difficulty. Underch. = underchallenging, ch. = challenging, overch. = overchallenging.

physiological parameters that showed significant differences between challenging and overchallenging conditions.

To better illustrate differences between difficulty levels, relative values of four physiological parameters are shown as graphs: *mean heart rate* (Fig. 2), *SCR frequency* (Fig. 3), *respiratory rate variability* (Fig. 4) and *final skin temperature* (Fig. 5).

4. Discussion

Since the goal of our study was to determine whether psychophysiological responses could differentiate between different levels of cognitive workload at different levels of physical workload, this section first focuses on the psychophysiological parameters that differentiate between different levels of cognitive workload at both levels of physical load, then examines the physiological effects of physical load in more detail.

4.1. *Differences between difficulty levels*

Many psychophysiological parameters showed significant differences between baseline and task (Table II), but only three parameters showed significant differences between difficulty levels regardless of the level of physical load (Table III). For both low and high physical load, *mean respiratory rate* (Fig. 3) showed a significant difference between underchallenging and the other two conditions while *final skin temperature* (Fig. 5) and *respiratory rate variability* (Fig. 4) showed a significant difference between the challenging and overchallenging conditions. Thus, a combination of respiration and skin temperature appears to be a robust method of cognitive workload estimation in physically demanding interaction with haptic robots.

Mean respiratory rate has been shown to be an indicator of cognitive workload and arousal.[19,20] Since self-reported arousal in the underchallenging condition is lower than in the other two conditions (Table I), it is not surprising that *mean respiratory rate* is also lower (Fig. 3). *Respiratory rate variability* is lower than baseline for all task conditions, but is lowest for the challenging condition (Fig. 4). A possible explanation is that it decreases as cognitive workload increases, but increases again as the challenge becomes too much to handle. Respiratory rate variability is known to decrease as a result of cognitive workload,[22] confirming part of this explanation.

Final skin temperature only significantly decreases from baseline in the overchallenging condition, not in the other conditions (Fig. 5). Thus, it may be a good indicator of when a subject is overworked. Previous studies have found decreases in skin temperature as a result of tension and anxiety,[24] supporting our hypothesis. However, other studies have found skin temperature to decrease as a result of cognitive workload.[23] If skin temperature decreases due to cognitive workload, it should also decrease during the challenging condition. One possibility is that a certain threshold of cognitive workload must be exceeded before skin temperature decreases.

4.2. *Influence of physical workload*

First of all, it should be noted that it is apparent from the mean absolute force measurements (Section 3.2) that subjects were not equally physically active in the three difficulty levels. Thus, we must be cautious when comparing physiological responses to different conditions. The change in a physiological response may not be caused primarily by changes in cognitive load, but instead by the increased physical load associated with task difficulty. However, the differences between low and high physical load were not significant in the case of respiration and skin temperature, suggesting that these responses are relatively robust to changes in physical load. This reinforces our suggestion that they are the most useful psychophysiological responses in haptic interaction. Skin temperature, in particular, seems the least affected by physical load. While neither *mean respiratory rate* (Fig. 3) and *respiratory rate variability* (Fig. 4) show significant differences between low and high physical load, non-significant differences are clearly visible in the corresponding figures. *Final skin temperature*, on the other hand, exhibits much smaller differences between low and high physical load (Fig. 5).

Only three physiological parameters showed a significant difference between low and high physical load: *mean heart rate*, *mean SCL* and *SCR frequency*. *Mean SCL*, despite its established connection to arousal and cognitive workload,[17,18] found no significant differences between difficulty levels at either level of physical load. Thus, it appears to be predominantly affected by physical activity. While *SCR frequency* showed a significant difference between underchallenging and the other two conditions in the case of low physical load, the difference was not quite significant in the case of high physical load (the p-value in question was 0.06 – Table III). This is most likely the effect of higher physical load in the underchallenging condition (see Section 3.2). If physical load had been the same for all the three difficulty levels, we expect that *SCR frequency* would have shown a significant difference between underchallenging and the other two conditions. Nonetheless, this suggests that *SCR frequency* is less robust than *mean respiratory rate* with regard to physical load.

Though heart rate has been used as a psychophysiological indicator in many studies, our results suggest that, in haptic interaction, it is primarily influenced by physical load. By far the greatest increase in *mean heart rate* was during the underchallenging condition (Fig. 2), where the exerted force was also the greatest. In all conditions with high physical load, *mean heart rate* was significantly higher than in the corresponding conditions with low physical load. Thus, it appears that the increase in heart rate due to physical workload can completely overshadow any psychological effects. Since the effects of physical workload on heart rate have been extensively studied, a possible solution would be to collect information about physical workload from sensors built into the haptic robot. This information could be used in conjunction with a physiological model to provide an estimate of the effects of physical load on heart rate. Such a model has already been developed for use in robot-assisted lower extremity rehabilitation,[25] so it should be possible to develop a similar model for the upper extremities.

5. Conclusions

We were able to demonstrate a significant psychophysiological difference between underchallenging, challenging and overchallenging tasks even in the presence of physical load. For both levels of physical load, *mean respiratory* rate showed a significant difference between the underchallenging condition and the other two conditions, while *respiratory rate variability* and *final skin temperature* showed a difference between challenging and overchallenging conditions. Skin conductance was also useful for low physical load, but was relatively vulnerable to the effects of high physical load. Although heart rate and heart rate variability did show differences between difficulty levels, those differences were most likely the result of physical load.

Our study has two important findings. First, it is clearly possible to use psychophysiological responses to differentiate between different levels of cognitive load in physically demanding haptic interaction. Second, it is best to use several psychophysiological measurements together. In our study, no single psychophysiological parameter was able to differentiate between the three difficulty levels for both levels of physical load. However, with two parameters it is possible to show differences between all the three difficulty levels. In our case, *mean respiratory rate* can differentiate between underchallenging and challenging while *final skin temperature* can differentiate between challenging and overchallenging.

Nonetheless, if physical load becomes too extreme, we expect that it would cause strong physiological responses that would completely obscure any physiological changes caused by changes in psychological state. This already partially happened with non-specific skin conductance responses in our study. 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. Such sensors are readily available in many haptic interfaces and would enable researchers to estimate the degree to which physiological measurements are affected by physical load and thus make corrections to estimations of psychological state.

Acknowledgements
The authors would like to thank the volunteers who participated in this study. The work was funded by the EU Information and Communication Technologies Collaborative Project MIMICS grant 215756.

References

1. J. J. Andreassi, *Psychophysiology: Human Behaviour and Physiological Response* (Lawrence Erlbaum Associates. London, UK, 2000).
2. M. Meehan, B. Insko, M. Whitton and F. P. Brooks, “Physiological measures of presence in stressful virtual environments,” *ACM T. Graphic.* **21**, 645–653 (2002).
3. M. Slater, C. Guger, G. Edlinger, R. Leeb, G. Pfurtscheller, A. Antley, M. Garau, A. Brogni, A. Steed and D. Friedman, “Analysis of physiological responses to a social situation in an immersive virtual environment,” *Presence: Teleop. Virt.* **15**, 553–569 (2006).
4. P. Rani, N. Sarkar, C. A. Smith and L. D. Kirby, “Anxiety detecting robotic system – towards implicit human–robot collaboration,” *Robotica* **22**, 85–95 (2004).
5. D. Kulić and E. Croft, “Physiological and subjective responses to articulated robot motion,” *Robotica* **25**, 13–27 (2007).
6. C. Liu, K. Conn, N. Sarkar and W. Stone, “Online affect detection and robot behavior adaptation for intervention of children with autism,” *IEEE Trans. Robot.* **24**, 883–896 (2008).
7. Y. J. Yao, Y. M. Chang, X. P. Xie, X. S. Cao, X. Q. Sun and Y. H. Wu, “Heart rate and respiration responses to real traffic pattern flight,” *Appl. Psychophysiol. Biofeedback* **33**, 203–209 (2008).
8. D. L. Roth, S. D. Bachtler and R. B. Fillingim, “Acute emotional and cardiovascular effects of stressful mental work during aerobic exercise,” *Psychophysiology* **27**, 694–701 (1990).
9. H. E. Webb, M. L. Weldy, E. C. Fabianke-Kadue, G. R. Orndorff, G. H. Kamimori and E. O. Acevedo, “Psychological stress during exercise: cardiorespiratory and hormonal responses,” *Eur. J. Appl. Physiol.* **104**, 973–981 (2008).
10. R. Q. van der Linde and P. Lammertse, “HapticMaster – a generic force controlled robot for human interaction,” *Ind. Robot* **30**, 515–524 (2003).
11. D. C. Fowles, M. J. Christie, R. Edelberg, W. W. Grings, D. T. Lykken and P. H. Venables, “Committee report: Publication recommendations for electrodermal measurements,” *Psychophysiology* **18**, 232–239 (1981).
12. M. M. Bradley and P. J. Lang, “Measuring emotion: the self-assessment manikin and the semantic differential,” *J. Behav. Ther. Exp. Psychiatr.* **25**, 49–59 (1994).
13. A. Mehrabian, “Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament,” *Curr. Psychol.* **14**, 261–292 (1996).
14. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, “Heart rate variability: Standards of measurement, physiological interpretation, and clinical use,” *Eur. Heart J.* **17**, 354–381 (1996).
15. J. A. Veltman and A. W. K. Gaillard, “Physiological workload reactions to increasing levels of task difficulty,” *Ergonomics* **41**, 656–669 (1998).
16. R. W. Backs, J. K. Lenneman, J. M. Wetzel and P. Green, “Cardiac measures of driver workload during simulated driving with and without visual occlusion,” *Hum. Factors* **45**, 525–538 (2003).
17. B. H. Detenber, R. F. Simons and G. G. Bennett, “Roll ‘em!: the effects of picture motion on emotional responses,” *J. Broadcast. Electron.* **42**, 113–127 (1998).
18. C. Collet, P. Averty and A. Dittmar, “Autonomic nervous system and subjective ratings of strain in air-traffic control,” *Appl. Ergon.* **40**, 23–32 (2009).
19. R. Nikula, “Psychological correlates of nonspecific skin conductance responses,” *Psychophysiology* **28**, 86–90 (1991).
20. A. Haarmann, W. Boucsein and F. Schaefer, “Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight,” *Appl. Ergon.* **40**, 1026–1040 (2009).
21. J. B. Brookings, G. F. Wilson and C. R. Swain, “Psychophysiological responses to changes in workload during simulated air traffic control,” *Biol. Psychol.* **42**, 361–377 (1996).
22. F. Boiten, “Component analysis of task-related respiratory patterns,” *Int. J. Psychophysiol.* **15**, 91–104 (1993).
23. M. Ohsuga, F. Shimono and H. Genno, “Assessment of phasic work stress using autonomic indices,” *Int. J. Psychophysiol.* **40**, 211–220 (2001).
24. B. C. Min, S. C. Chung, S. J. Park, C. J. Kim, M.-K. Sim and K. Sakamoto, “Autonomic responses of young passengers contingent to the speed and driving mode of a vehicle,” *Int. J. Ind. Ergonom.* **29**, 187–198 (2002).
25. A. C. Koenig, L. Somaini, M. Pulfer, T. Holenstein, X. Omlin, M. Wieser and R. Riener, “Model-Based Heart Rate Prediction During Lokomat Walking,” *Proceedings of the 31st Annual International Conference of IEEE EMBC*, Minneapolis (2009).