

UNIVERSITÀ DEGLI STUDI DI UDINE

DIPARTIMENTO DI MATEMATICA E INFORMATICA

DOTTORATO DI RICERCA IN INFORMATICA

PH.D. THESIS

**Stress Detection with Physiological Sensors for
Evaluating Interactive Systems and Building
Relaxation Training Applications**

CANDIDATE:

Riccardo Sioni

SUPERVISOR:

Luca Chittaro

Academic Year 2013–2014

Author's e-mail: riccardo.sioni@uniud.it

Author's address:

Dipartimento di Matematica e Informatica
Università degli Studi di Udine
Via delle Scienze, 206
33100 Udine
Italia

Contents

Introduction	xi
1 Psychological and Physiological Background	1
1.1 Emotions	1
1.1.1 What are emotions?	1
1.1.2 Emotions representation in the literature	2
1.2 Stress and Anxiety	3
1.2.1 General definition of stress and anxiety	3
1.3 Physiological Measurements for Stress	4
1.3.1 Electrodermal activity	4
1.3.2 Facial muscle activity and eye–blink startle response	6
1.3.3 Cardiovascular system activity	8
1.3.4 Respiratory system activity	10
1.3.5 Brain activity	11
2 Affective State Detection and Its Applications to Stress and Anxiety Management	13
2.1 Algorithms and Techniques for Automatic Stress Detection	13
2.1.1 Fuzzy logic	13
2.1.2 Bayesian networks and influence diagrams	14
2.1.3 Hidden Markov models	14
2.1.4 Naïve Bayes classifiers	14
2.1.5 Support vector machines and linear discriminant analysis	14
2.1.6 K–nearest neighbor algorithm	15
2.1.7 Decision tree classifiers	15
2.2 Automatic Emotion Detection and Affective Computing Systems	15
2.3 Applications of Affective Computing to Stress and Anxiety Management	18
2.3.1 Training and biofeedback	18
2.3.2 Performance and health monitoring during work	20
2.3.3 Exposure therapy	20
3 Physiology–Based Evaluation of Interaction Devices	21
3.1 Introduction	21
3.2 Laser Pointer–Style Interaction in the Literature	22
3.2.1 Techniques based on ray casting	23
3.2.2 User studies	23
3.3 The Considered Interaction Techniques	24
3.3.1 The Wii Remote	24
3.3.2 Navigating the virtual environment	26
3.3.3 Pointing at and selecting an object	26
3.3.4 Manipulating the selected object	27
3.4 Experimental Evaluation	28
3.4.1 Participants	28
3.4.2 Materials	28
3.4.3 Task	29
3.4.4 Procedure	30
3.4.5 Data collection	32
3.5 Results	33

3.5.1	Task completion time	33
3.5.2	Navigation time, coarse manipulation time, and fine manipulation time	33
3.5.3	Circulatory and respiratory activity and temperature measurements	37
3.5.4	Mean muscle activity	39
3.5.5	Total muscle activity	40
3.5.6	EMG power spectrum mean frequency	41
3.5.7	EMG gradients	42
3.5.8	Linear regression slopes of the mean values of the EMG power spectrum mean frequency	44
3.5.9	Ranked choice questionnaire	44
3.6	Discussion of the Results	44
3.6.1	User performance	44
3.6.2	Muscle activity	46
3.6.3	Other physiological measurements	47
3.6.4	Subjective preferences	47
4	Physiology–Based Evaluation of Video Game User Experience	49
4.1	Introduction	49
4.2	Related Work	50
4.2.1	Physiological measurements and video game UX	50
4.2.2	Effects of violent video games	51
4.2.3	Effects of violent video games on aggressive feelings, thoughts and behaviors	51
4.2.4	Effects of violent video games on desensitization to violence	53
4.3	The Considered Video Games	53
4.4	Experimental Evaluation	55
4.4.1	Participants	55
4.4.2	Materials	56
4.4.3	Procedure	56
4.4.4	Measures	58
4.4.5	Possible outcomes	58
4.5	Results	59
4.5.1	Facial EMG	59
4.5.2	HR, BVPA, and SCL	59
4.5.3	AG scores	60
4.5.4	Correlation analysis	60
4.6	Discussion of the Results	60
4.6.1	Players’ affect during game play	60
4.6.2	Desensitization to violence	63
4.6.3	Correlation analysis	63
4.6.4	Limitations of the study	63
5	Exploring Eye–Blink Startle Response as a Physiological Measure for Affective Computing	65
5.1	Introduction	65
5.2	Eye–Blink Startle Response and Virtual Environments	66
5.3	The Selected Startle Stimuli	66
5.4	The Considered Virtual Environment	67
5.5	Experimental Evaluation	68
5.5.1	Participants	68
5.5.2	Materials	68
5.5.3	Procedure	69
5.6	Results	69
5.6.1	Data reduction and analysis	69
5.6.2	Results	70
5.7	Discussion of the Results	71

6	Physiology–Based Evaluation of a Mobile Breathing Training Application	73
6.1	Introduction	73
6.2	Related Work	74
6.3	Motivations and Goals	75
6.4	The Considered Designs	76
6.5	Experimental Evaluation	77
6.5.1	Materials	77
6.5.2	Participants	78
6.5.3	Procedure	78
6.5.4	Measures	78
6.6	Results	79
6.7	Discussion of the Results	79
7	The Proposed Affective State Detection System: ACME	83
7.1	Introduction	83
7.2	ACME Architecture	84
7.2.1	Input data	84
7.2.2	Raw data extraction and real–time analysis module	85
7.2.3	Data normalization module	85
7.2.4	Arousal/valence fuzzy system module and stress fuzzy system module	86
7.2.5	Communication module	86
7.3	ACME User Interface	86
7.4	ACME Preliminary Evaluation	88
7.4.1	Pilot study with a training application for evacuation during a fire emergency	89
7.4.2	Preliminary evaluation of ACME	91
8	Physiology–Based Evaluation of a Biofeedback–Controlled Game for Relaxation Training	93
8.1	Introduction	93
8.2	Related Work	94
8.3	The Biofeedback Game	95
8.3.1	First level of the game	96
8.3.2	Users’ stress detection	97
8.3.3	The placebo condition algorithm	100
8.4	Experimental Evaluation	100
8.4.1	Participants	100
8.4.2	Materials	100
8.4.3	Procedure	101
8.4.4	Hypothesis	101
8.5	Results	102
8.6	Discussion of the Results	103
	Conclusions	105
C.1	Physiology–Based Evaluation of Interaction Devices	105
C.2	Physiology–Based Evaluation of Video Game User Experience	106
C.3	Novel Approaches for Stress and Anxiety Detection	107
C.4	Physiology–Based Evaluation of a Mobile Breathing Training Application	107
C.5	Physiology for Building and Evaluating a Biofeedback–Based Relaxation Training Application	108
	Bibliography	111

List of Figures

1.1	The circumplex model of affect, as presented in (Russell, 1980).	3
1.2	A commercial EDA sensor.	5
1.3	EDA signal characteristics. From (Dawson et al., 2007)	5
1.4	Screenshot from Ledalab, a Matlab-based software to decompose the EDA signal into SCR and SCL based on (Benedek and Kaernbach, 2010). The tonic (gray) and phasic (blue) components of the EDA signal are shown. The red vertical lines represent the discrete stimuli.	6
1.5	Suggested electrode placements for facial EMG measurements, from (Tassinari et al., 2007).	7
1.6	Various representations of the EMG signals, from (Tassinari et al., 2007). From top to bottom, (i) raw EMG signal; (ii) and (iii) rectified signal; (iv) smoothed EMG; (v) integrated EMG.	8
1.7	A commercial photoplethysmograph.	9
1.8	ECG and BVP signals.	9
1.9	On the left, a respiratory sensor; on the right, an example of respiratory signal.	10
1.10	Electrodes position and labels in the International 10–20 System.	12
2.1	A screenshot from the biofeedback game presented in (Gorini et al., 2010). A reduction in heart rate reduces fire intensity, and vice versa.	19
3.1	A commercial laser pointer.	22
3.2	(a) The Nunchuck (on the left) connected to the Wii Remote (on the right); (b) the Sensor Bar used in the study (the two red circles highlight the position of the two IR LED arrays).	25
3.3	The controls for the two interaction techniques. <i>a</i> : navigation; <i>b</i> : viewpoint orientation, object selection and coarse manipulation; <i>c</i> : manipulation type; <i>d</i> : fine manipulation.	26
3.4	From left to right, the three cursor icons associated respectively to translation, rotation and size manipulation.	27
3.5	Position of the seven sensors. The silhouette represents the back of a human body.	29
3.6	An object to be arranged (on the right) and the corresponding target object (on the left). In this example, the object to be arranged is currently selected (bounding box highlighted).	29
3.7	(a) User’s viewpoint at the beginning of the task; (b) the user has matched the “K” target, and is arranging the “Z” object.	30
3.8	A participant performing a task with W&N in sitting position.	31
3.9	An example of room map given to participants. The participant’s starting position as well as viewing direction are represented by the central dot and arrow. Colored letters show the position of the objects to be manipulated (green and blue letters, at the top of the map) and the targets (red letters, at the bottom). Note that the targets are initially outside the users’ field of view.	31
3.10	Mean completion time. Error bars indicate standard error of the mean.	34
3.11	(a) Mean navigation time in seconds and (b) mean navigation time expressed as a percentage of the sum of navigation, coarse and fine manipulation times. Error bars indicate standard error of the mean.	34
3.12	(a) Mean coarse manipulation time in seconds and (b) mean coarse manipulation time expressed as a percentage of the sum of navigation, coarse and fine manipulation times. Error bars indicate standard error of the mean.	35
3.13	(a) Mean fine manipulation time in seconds and (b) mean fine manipulation time expressed as a percentage of the sum of navigation, coarse and fine manipulation times. Error bars indicate standard error of the mean.	35

3.14	Mean time, in percentage of the sum of navigation time, coarse manipulation time and fine manipulation time, spent on coarse and fine translation, rotation and size manipulation. Error bars indicate standard error of the mean.	36
3.15	Change of BVPA (chart with reversed scale) and HR with respect to baseline values. Error bars indicate standard error of the mean.	37
3.16	Change of respiration amplitude and respiration frequency with respect to baseline values. Error bars indicate standard error of the mean.	38
3.17	Change of peripheral temperature (chart with reversed scale) with respect to baseline values. Error bars indicate standard error of the mean.	38
3.18	Change of mean activity for left and right extensor digitorum communis with respect to baseline values. Error bars indicate standard error of the mean.	39
3.19	Change of mean activity for right superior trapezius and right biceps brachii with respect to baseline values. Error bars indicate standard error of the mean.	39
3.20	Change of IEMG data for left and right extensor digitorum communis with respect to baseline values. Error bars indicate standard error of the mean.	40
3.21	Change of IEMG data for right superior trapezius and right biceps brachii with respect to baseline values. Error bars indicate standard error of the mean.	40
3.22	Change of mean power spectrum frequency for the four muscles (charts with reversed scale) with respect to baseline values. Error bars indicate standard error of the mean.	41
3.23	Change of EMG gradient data for the four muscles. Error bars indicate standard error of the mean.	42
3.24	Linear regression slope of left and right extensor digitorum communis (charts with reversed scale), right superior trapezius and right biceps brachii. Error bars indicate standard error of the mean.	43
3.25	Mean values for the eight items of the questionnaire. Ranks range from 1 (best) to 4 (worst). Error bars indicate standard error of the mean.	45
4.1	Screenshots of the violent Whac-a-Mole game: (a) the wood ants are walking on the chair; (b) two ants are squashed. Note the wood shavings on the floor.	54
4.2	<i>Game Over</i> screen with the broken chair.	54
4.3	Screenshots of the non-violent Whac-a-Mole game: (a) two purple rectangles appear; (b) the user has touched the two rectangles and a new rectangle is appearing.	55
4.4	Position of PPG end EDA sensors during the experiment.	56
4.5	The AG template used in our study.	57
4.6	Location of left and right CS and ZM.	57
4.7	Mean CS and mean ZM activity recorded during the gaming session, after baseline subtraction. Error bars indicate standard error of the mean.	60
4.8	Mean CS and mean ZM activity recorded while watching the two videos, after baseline subtraction. Error bars indicate standard error of the mean.	61
4.9	Mean HR, BVPA, and SCL values after baseline subtraction, recorded during the gaming session as well as while participants were watching the two violent videos. Error bars indicate standard error of the mean.	62
4.10	Scatterplot of VGQ and AQ scores. The regression line is also outlined.	62
5.1	A screenshot of the VE employed in the study showing a corridor full of smoke and corpses lying on the floor.	67
5.2	A screenshot of the VE showing the red aura.	68
5.3	The two black dots pinpoint the location of EMG electrodes for EBSR measures.	69
5.4	Filtered EMG signal showing an EBSR to a stimulus presented in correspondence to the dot-dashed vertical line.	70
5.5	Mean magnitude of EBSRs to the white noise and the explosion sound. Error bars indicate standard error of the mean.	70

6.1	Two screenshots from the Sphere design during inhalation (a) and exhalation (b) phases, in which the sphere is green and red respectively.	76
6.2	Two screenshots from the Wave design during the exhalation (a) and inhalation (b) phases.	77
6.3	Mean values of power of the recommended frequency band (a), respiratory SNR (b), perceived effectiveness (c), perceived relaxation effectiveness (d), SCL (e), HR (f). Error bars indicate standard error of the mean. a.u. indicates the arbitrary units provided by the elastic girth sensor to measure its tension.	80
6.4	Mean values of subjective preference. A lower value indicates higher preference. Error bars indicate standard error of the mean.	81
7.1	The ACME architecture	84
7.2	The fuzzy sets employed by ACME to translate normalized physiological data into arousal and valence indices.	87
7.3	The fuzzy sets employed by ACME to translate arousal and valence indices into a stress index.	87
7.4	The ACME interface.	88
7.5	One of the stressful stimuli: fire is blocking one of the escape routes.	89
7.6	The moment when the user, sat on the wagon, see the explosion.	90
7.7	The predefined safe path, shown as a red line over a top view of the train station. The dotted portion of the line indicates that the user has to take the subway.	90
7.8	Two screens of the 3D VE employed for the training session. (a) A maze designed to train users to employ W, A, S, and D keyboard keys to move inside the 3D VE; (b) a checkpoint along the training path.	91
7.9	Arousal, valence, and stress indices recorded from the Participant no. 2 during the experiment. The first red marker points out the end of the baseline session; the second red marker points out the moment when the first explosion occurs during the play session, i.e., when the participant was sit on the wagon. Note how the arousal and stress indices sharply increase in correspondence of the second marker.	92
7.10	Mean values of arousal index data. Error bars indicate standard error of the mean.	92
8.1	Screenshots of the virtual character in its five different states.	98
8.2	Screenshots of the virtual character raising its hands and punching the table.	99
8.3	An example of use of the EDA-only algorithm.	99
8.4	Mean values of the perceived quality of the biofeedback score. Error bars indicate standard error of the mean.	102
8.5	Mean values of the difficulty of relaxation training score. Error bars indicate standard error of the mean.	102

List of Tables

2.1	Some of the ACSs presented in the literature since 2001 that use physiological sensors to perform emotion detection.	16
3.1	Time spent by users in viewpoint orientation, split in percentage between usage of mouse and Wii Remote pointing, and usage of keyboard buttons and Nunchuck joystick. The third column provides the average difference of the two percentages.	37

Introduction

The expression *affective computing* denotes computing that relates to, arises from or influences emotions (Picard, 1997). *Affective computing systems* (ACSs) must be able to quantify users' affective state, which influences the system response accordingly: when it is integrated into a computer application, the ACS output is employed to change the application state with a main goal of altering users' emotions. For example, an ACS applied to a *first-person shooter* (FPS) video game could increase or decrease the number of enemies on screen during a play session to modulate user engagement and reduce users' boredom or frustration. An ACS integrated into a VE-based training application (i.e., an application that, with the use of immersive virtual environments, allows the acquisition of knowledge, skills and competencies through direct or indirect experience) can modulate the intensity of stressful stimuli to increase trainees' habituation to stressful situations and the ability to maintain a high level of performance.

Psychologists are studying the nature of emotions since the 19th century (James, 1884; Schachter and Singer, 1962), without reaching an universally accepted definition of what are emotions and how emotions are generated. However, over a century of research shows that emotions and bodily functions are related. It is not surprising, therefore, that many studies in the affective computing literature employ physiological data such as cardiovascular, muscular, and brain activity to detect users' emotions. Other instruments could be used to assess users' affective state, such as questionnaires and scales. However, they cannot be administered to participants without interrupting the experiment (and thus potentially affecting the participants' emotions); furthermore, the intrinsic ambiguity in describing emotion through written words could undermine the reliability of such instruments.

Despite all the research work, the design of an ACS remains a very difficult task, for two reasons. First, as reported by Fairclough (2009), there are still issues that have a critical impact on the development and evaluation of ACSs: for example, instances of unique one-to-one relationship between a physiological measure and a psychological state, which are ideal for ACSs, are very rare in the psychophysiological literature relative to one-to-many, many-to-one, and many-to-many relationships. According to Fairclough, these relationships represent physiological links that are neither exclusive nor uncontaminated. Furthermore, every person is physiologically different from the others, which makes it difficult to develop an ACS equally accurate for every possible user.

Second, it is not trivial to appropriately adjust the application state to influence users' emotions in the desired way. As noted by Muller (2004), emotions can be experienced for a variety of reasons: for example, while playing a FPS, stress can be elicited by a malfunctioning controller, a difficulty level perceived as too high, an increase of the room temperature, etc. In those cases, even if the ACS has correctly detected the user's affective state, it may not be able to identify the reason for the high stress level, and thus may incorrectly modify the application state. Also, it must be considered that application state changes may be required to be substantial to have effects on users' emotional state that can be detected and further evaluated by the ACS. Indeed, Ward and Marsden (2004) observed that bodily responses are evident in the case of major interface events and gross failures of usability, while they had not always been able to find consistent statistically significant events related to weaker stimuli. For some of the applications that can be considered for ACS integration, the design and implementation of state changes of the required amplitude may be very difficult, and must be planned carefully. For example, one can hypothesize that, in order to reach a sufficiently low stress level in users, the number of enemies in a FPS must be drastically reduced. This is an approach that is commonly employed in video games based on affective computing (e.g., Nacke and Lindley, 2009). If enemies are currently thousands, it is hard to justify their sudden disappearance consistently with the gaming narrative; at the same time, if the enemies are few, the game designer must provide other means to the ACS to lower the stress level, like modifying the background music, or the environment appearance (e.g., Riva et al., 2007).

The approaches of some of these ACSs are indeed criticized by Muller (2004): the issues in emotion quantification and in the modification of the application state might have had a role into the less than optimal results obtained by some of the ACSs proposed in the literature (Kim et al., 2004; Rani et al., 2007).

The goal of this thesis is to employ physiology to deal with some topics that have been scarcely explored in the HCI literature, and to advance the state of the art in affective computing by building bridges among HCI, psychology and physiology. In particular, the thesis aims to:

- Highlight the importance of exploiting physiology to evaluate novel interaction devices and techniques;
- Propose new methods for exploiting physiology to build and evaluate applications for relaxation training, a research topic in the field of affective computing;
- Extend the current body of knowledge of affective computing by proposing new approaches for improving stress detection and new methodologies for the evaluation of ACSs.

In Chapter 1, we report some of the various definitions of emotions presented in the literature, with particular focus on stress and anxiety. Also, we describe how emotions are typically represented in the psychology and psychophysiology research, as well as the most common physiological measurements employed to assess stress and anxiety. Chapter 2 describes various algorithms and techniques employed in the literature for stress detection, outlines the current state of the art of emotion detection systems and ACSs, and describes possible applications of ACSs, such as in the field of training. We then deal with the use of traditional physiology measurements in user evaluations. In a first study, described in Chapter 3, we employ users' physiology to assess the ergonomics of innovative devices to interact with 3D objects in a *virtual environment* (VE). A second user evaluation, described in Chapter 4, uses physiology to measure users' emotional reactions in terms of UX afforded by two versions of a mobile video game. In this study, we employ well-known physiological signals, described in Chapter 1, to evaluate the physiological responses to violent content in the form of aggression towards non-human animals. To improve the current body of knowledge of affective computing literature, in Chapter 5 we then turn our attention from the physiological signals employed in Chapter 4 to *eye-blink startle response* (EBSR), a physiological signal well known in the psychophysiology literature, but rarely employed in affective computing studies. In Chapter 6, we focus on *training*, introduced in Chapter 2, and present a study that extends the use of physiology to the evaluation of breathing training applications for relaxation. With the experience gained on emotion detection from the research activity described in previous chapters, we introduce the *Affective Computing MachinE* (ACME), an ACS for automatic stress detection, in Chapter 7. ACME exploits the physiological signals that, according to the studies presented in previous chapters, resulted the most suitable for performing stress detection. In Chapter 7 we also describe a preliminary evaluation of the system. Chapter 8 combines all the topics discussed in the previous chapters. We propose a game designed to support relaxation training that employs realistic 3D graphics to provide players with embodied feedback. Also, we focus on methodological aspects of ACS evaluations, in particular on the use of a placebo condition to better evaluate the effectiveness of the proposed game, which was tested both with ACME and a second algorithm based on a single physiological sensor to assess trainees' stress levels. Finally, we conclude the thesis by discussing future directions for our research.

1

Psychological and Physiological Background

In this chapter, we introduce two key elements of this thesis: *emotions* and *physiology*. In Section 1.1, we report some of the various definitions of emotions that have been proposed in the literature since the late XVIII century. We focus on the relationship between emotions and physiology, and how emotions are typically represented in the literature. In Section 1.2, we focus our attention on stress and anxiety. In Section 1.3, we describe in detail the most common physiological measurements employed to assess stress and anxiety. In particular, we describe how physiological data is measured and analyzed, what instruments are employed, and the relationship between each physiological signal and a person's affective state.

1.1 Emotions

1.1.1 What are emotions?

As Bradley and Lang (2007) emphasize, there are as many definitions of emotions as many emotion investigators in the literature, but most researchers agree that in emotional situations, the body acts, i.e., “the heart pounds, flutters, stops and drops; palms sweat; muscles tense and relax; blood boils; faces blush, flush, frown, and smile”. Correlations between affective state and bodily functions were already theorized more than a century ago by James in his work about psychology and emotions (1884; 1890), and with the contribution of Lange (1887). In the original *James–Lange theory*, emotions were defined as the interpretation of the experienced bodily expression: for example, if one sees a big scary dog barking at her, her heart begins to race. Noticing her heart race, her brain figures out that she is experiencing fear.

A few years later, Cannon (1927) criticized the James–Lange theory, and proposed a new theory in collaboration with one of his doctoral students, Bard. The *Cannon–Bard theory of emotion* states that visceral responses do not generate emotions. Emotions, instead, are activated by *thalamic processes*, i.e., discharges from neurons in the thalamic region. These discharges also have an immediate impact on contiguous neurons innervating muscles and viscera, causing the bodily responses related to the emotion expressions. Other theories similar to the Cannon–Bard one were later proposed. In general, these theories differ from the Cannon–Bard theory on the area responsible for the emotion expressions; for example, the *Papez–MacLean theory* (1937) focuses on the role of the *limbic system*, and in particular on the activity of the *hippocampus*.

Schachter and Singer (1962) proposed a theory (*Schachter–Singer theory*, or *two–factor theory of emotions*) that focuses on the role of cognition in the *interpretation* of high–arousal physiological states (i.e., states in which the person is “awake” and highly reactive to stimuli) as emotions. The Schachter–Singer theory states that emotions are the result of a person's physiological responses, and (s)he can successfully interpret (i.e., explain) these responses thanks to her cognitive appraisal of the environmental cues readily available. While this theory, in the 1960s, was considered as fundamental as James–Lange and Cannon–Bard theories, the Schachter and Singer study was later heavily criticized, mainly because of the methodological and conceptual flaws found by many psychophysicists. Nevertheless, at that time the

cognitive models of affect continued to be predominant. Lazarus (1968) defines emotions as disturbances caused by an intentional cognitive appraisal of an event: for example, after one perceives the presence of a bear in her kitchen (the event), she actively assesses the dangerous situation (the cognitive appraisal), and her physiology changes (e.g., her heartbeat increases) to prepare her to run away as fast as possible.

It is only in the 1980s, with the seminal work by Ekman et al. (1983) and Scherer (1984) that a renewed interest could be observed in the relation between emotions and physiological responses, originally proposed by James. Ekman et al. (1983) observed that different basic emotions (anger, fear, sadness, happiness, disgust and surprise) are related to different physiological responses; in particular, Ekman and Friesen (1971) discovered that these six basic emotions can be universally associated to unique facial expressions. The *Component-Process Model* (CPM) by Scherer (1984) describes emotion as a series of interrelated adaptive changes in several organismic subsystems following antecedent events evaluated to be of major relevance to an organism's goals. As reported in (Friedman, 2010), this strong relation between emotions and bodily functions is supported by numerous studies. The results of one meta-analysis (Cacioppo et al., 2000) suggest that positive and negative emotions can be differentiated on the base of physiological responses. Furthermore, some studies reported evidence that *interoception* (the awareness of one's internal bodily states) and emotional experience share the same functional neural architecture, i.e., both activate a circumscribed area in the insular cortex, a portion of the cerebral cortex (Zaki et al., 2012).

1.1.2 Emotions representation in the literature

As reported in the previous section, there is no unambiguous definition of what emotions are, or what originates them. It is not surprising, therefore, that there are different ways of representing emotions in the literature, every one with its advantages and disadvantages.

Bradley and Lang (2007) describe various studies which employ a *biphasic* (also known as *dimensional*) representation of emotions, differentiating emotional events between good or bad, appetitive or aversive, agreeable or disagreeable, positive or negative, pleasant or unpleasant, hospitable or inhospitable, etc. The bipolar nature of emotions seems to be strongly related to animal motivations which, like emotions in humans, describe animal behavior. As reported in (Bradley and Lang, 2007), the studies on animal behavior agree that action is controlled by two basic parameters, direction (*approach* vs. *avoid*) and intensity. Stimuli that promote survival elicit more or less intense approach behaviors, while threatening stimuli elicit escape, avoidance or withdrawal behaviors.

The *circumplex model* (Russell, 1980) is a well-known representation of emotions related to the biphasic representation. It locates emotions on a two-dimensional plane on the basis of their valence (positive or negative) and degree of arousal (low or high) (Figure 1.1). Stress and anxiety can be represented as high-arousal, negatively-valenced emotion, and located in the upper left quadrant of the circumplex model (Russell, 1980).

Emotions can be described in terms of categories as discrete affective states, such as joy, frustration, anger, fear and so on. In the literature, various emotion sets have been evaluated. Ekman (1972), on the base of his observations around the world, defined six basic emotions universal to human culture: anger, disgust, fear, happiness, sadness and surprise.

Both biphasic and discrete representations of emotions have been employed in emotion detection (e.g., Picard et al., 2001; Mandryk and Atkins, 2007). However, Bradley and Lang (2007) reported that already in 1958 researchers casted doubts on the discrete categorization of emotions. Also, the thorough review by Mauss and Robinson (2009) tends to support the dimensional perspective, noticing how dimensions appear to capture the lion's share of variance of emotional responses. One of the limits of the discrete representation of emotions is the necessity to create an absolutely unambiguous association between an emotion and a particular word (such as *joy*) to discern it from similar words (e.g., *happiness*). Furthermore, the available data are incompatible with the notion that discrete emotional states are categorically different from one another (Mauss and Robinson, 2009). Ambiguity in describing emotions is particularly detrimental when using subjective evaluations (i.e., questionnaires), because it could make data incomparable among subjects.

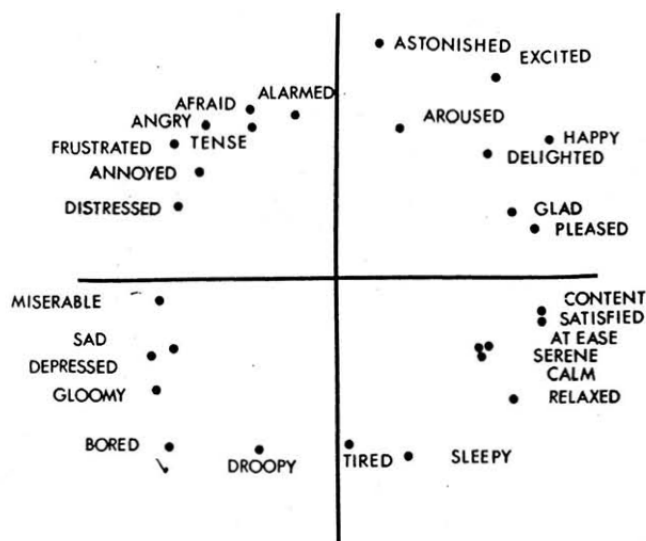


Figure 1.1: The circumplex model of affect, as presented in (Russell, 1980).

1.2 Stress and Anxiety

1.2.1 General definition of stress and anxiety

In common language, the term *stress* is often used with a negative meaning to represent strain caused, for example, by physical or psychological pressures in the workplace or at school. Indeed, Lazarus (1991) defines physiological stress as the result of the comparison between the power of the environment demands (called stressor or load) and the psychological resources of the person to manage these demands. In psychology (e.g., Öhman, 2008), however, stress is not always a bad thing. Some researchers (e.g., Kopin et al., 1988) make a distinction between *eustress* and *distress*, where *eustress* represents positive stress, such as excitement, or a stress leading to a state which is more beneficial to the organism.

Stress, however, is not an emotion; instead, it is the symptom of particular emotive states. Lazarus and Folkman (1984) defined stress as a particular relationship between the person and the environment, appraised by the person as taxing or exceeding her resources and endangering her well-being. From this point of view, stress can be seen as a defensive process used to protect oneself from potential injury or death. It is not surprising, therefore, that Öhman (2008) related stress to the capacity to adapt and respond to various circumstances, considering it as one of the pivotal elements of Darwin's theory of evolution.

In the psychology literature, *anxiety* is often used to denote a negatively-valenced emotion related to stress. In the Diagnostic and Statistical Manual of Mental Disorders, the term *anxiety* is used to characterize apprehensive anticipation of future danger or misfortune accompanied by a feeling of dysphoria (i.e., an intense experience of depression and discontent) or somatic symptoms of tension (American Psychiatric Association, 2000). Anxiety can be regarded both as an emotional state, evoked in a particular context and having a limited duration, and as a personality trait, characterizing individuals across time and situations (Öhman, 2008).

Other emotional states related to stress and anxiety are *fear*, *worry* and *rumination*. Fear denotes dread of impending and identifiable disaster and an intense urge to defend oneself (Öhman, 2008). Fear is related to coping behavior; when coping attempts fail, fear is turned into anxiety. Delgado et al. (2009) refers to fear and anxiety as *cued* and *non-cued* fear respectively, giving more prominence to the concreteness of the disaster. Worry has been conceptualized as a chain of thoughts and images, negatively affect laden and relatively uncontrollable, that promotes mental attempts to avoid anticipation of potential threats (Borkovec et al., 1991; Borkovec, 2002). From this perspective, worry can be described as a state of anticipatory anxiety. Rumination is a construct similar to worry, and is defined as thoughts or behaviors that focus attention towards depressive symptoms or their consequences, rather than towards neutral or pleasant topics

(Barlow, 1991). Worry and rumination characterize what studies cited by Sass et al. (2010) call anxious apprehension. *Anxious arousal*, on the other hand, is characterized by somatic tension and physiological arousal (Sass et al., 2010).

1.3 Physiological Measurements for Stress

The physiological responses which affective computing focuses on are controlled by the *autonomic nervous system* (ANS). The ANS is part of the nervous system, and controls important bodily activities, including digestion, body temperature, blood pressure, and many aspects of emotional behavior (Andreassi, 2007) which are under the level of consciousness. The ANS involves innervation of three types of cell: smooth muscle cells, which can be found for example within the walls of blood vessels and arteries; cardiac muscle cells; and glandular (secretory) cells. The ANS is subdivided into the *sympathetic nervous system* (SNS) and the *parasympathetic nervous system* (PNS). The SNS controls those activities that are mobilized during emergency and stress situations, which are represented by the *fight-or-flight* response. The activities under control of the PNS are the basic functions of rest, repair and relaxation of the body and restoration of energy stores (*rest-and-digest* activity) (Andreassi, 2007). Within the PNS, the *vagal activity* (i.e., the activity of the vagus nerve) is particularly important, because it contrasts sympathetic arousal by lowering heart rate, blood pressure, or both.

Mammals display various behavioral responses to state anxiety and fear (e.g., Lang et al., 2000; Vila et al., 2007; Sylvers et al., 2011). Lang et al. functionally organized them in two classes, (i) defensive immobility, in which the organism is passive but “primed” to respond actively to further stimulation, and (ii) defensive actions, which are more or less direct responses to *nociception* (i.e., the neural processes of encoding and processing harmful or unpleasant stimuli) or imminent attack. Responses such as *freezing* (becoming suddenly motionless or paralyzed), *fainting* (losing consciousness), *fear bradichardia* (slowed heartbeat), and *hyper-attentiveness* (increased alertness and decreased attention-focusing ability) belong to the first class; *startle* (which causes, among others effects, the contraction of muscles and blinking) and the *fight-or-flight* responses belong to the second class.

In the following, most of the physiological measurements for stress and anxiety used in the literature are presented.

1.3.1 Electrodermal activity

Studied since the 1880s (Boucsein, 2006), *electrodermal activity* (EDA) measures the change in the electrical conductivity of the skin surface (Andreassi, 2007). Changes in skin conductance can be produced by various physical and emotional stimuli that trigger variations in the eccrine sweat gland activity which, unlike many other bodily functions, is controlled exclusively by the SNS (Boucsein, 2006), making EDA a perfect physiological signal for arousal and stress measurement. In particular, as reported by Boucsein (2006), EDA is regarded as a sensitive and valid indicator for the lower arousal range, reflecting small, mostly cognitively determined, variations in arousal.

In situations that evoke high levels of arousal, an increase in the activity of the sweat glands can be observed (Boucsein, 2006), leading to what is called *emotional sweating*, which is observed mainly on palmar and plantar sites. Indeed, EDA is usually measured on the palms of the hands or on the fingers, with a current (imperceptible to users) that flows between a pair of electrodes (Figure 1.2) applied to two adjacent fingers (Andreassi, 2007) (either forefinger and middle finger, or ring finger and little finger). There is normally no need for a pretreatment (i.e., skin cleaning) of the sites used for EDA recording.

The EDA signal, measured in *microSiemens* (μS), can be split into two different components. *skin conductance level* (SCL) refers to the tonic component of the EDA signal, i.e., the electrical conductivity at a given point in time. *skin conductance response* (SCR) is the phasic component of EDA, i.e., the amount of change in EDA that occurs in response to a given stimulus (Andreassi, 2007). SCRs are considered an indicator of localized phasic arousal processes, while SCL was suggested to be useful for measuring more generalized arousal (Boucsein, 2006). As reported by Boucsein (2006), SCL and SCR have been successfully used in the literature as stress indicators during the presentation of all kinds of stressful stimuli, and constitute a valuable tool to be used in the area of stress research.



Figure 1.2: A commercial EDA sensor.

SCL values are generally in the $2\text{--}20\ \mu\text{S}$ range, while SCR amplitudes can vary from 0.1 to $5\ \mu\text{S}$ (Andreassi, 2007; Dawson et al., 2007). The typical EDA signal is characterized by slow and gradual changes (see Figure 1.3): a peak in the EDA signal starts $1\text{--}3$ s after the stimulus, and has a duration of about 4 s (Dawson et al., 2007). The slow-changing nature of EDA makes electrical interferences and artifacts (caused, for example, by user movements) easily detectable and removable. SCL and SCR can be obtained from the EDA signal with various procedures like the ones described in (Lim et al., 1997) and in (Benedek and Kaernbach, 2010), which *Ledalab* (Golz and Kaernbach, 2013), a Matlab-based software which is able to decompose previously recorded EDA data (Figure 1.4), relies upon. Once SCL and SCR signals are obtained, many features can be extracted, the two most common being:

- Mean SCL value, calculated over a given time interval;
- Mean SCR amplitude, i.e., the mean value of the SCRs peaks.

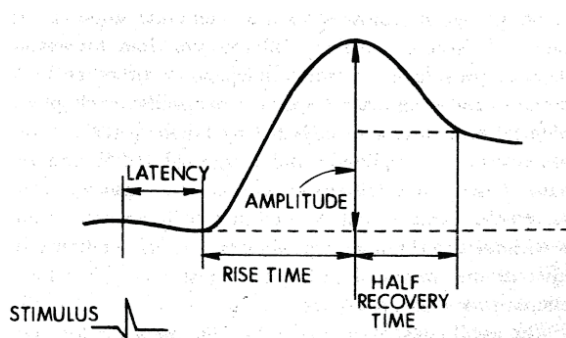


Figure 1.3: EDA signal characteristics. From (Dawson et al., 2007)

Other features are described in (Dawson et al., 2007). Usually, to consider a single SCR a non-zero response, an amplitude criterion is applied, for which the EDA increase must surpass $0.05\ \mu\text{S}$ (Boucsein, 2006). The mean SCR amplitude is calculated from the non-zero responses, while the mean SCR magnitude is calculated also considering the zero responses. Dawson et al. (2007) suggest to calculate the two measures separately, since there are arguments for and against both amplitude and magnitude: a magnitude measure can create the impression that the response size is changing, when in fact it is the response frequency that is changing; on the other hand, a subject may have a mean SCR amplitude of $0.5\ \mu\text{S}$ by having ten $0.5\ \mu\text{S}$ responses to ten different stimuli, or by having only one non-zero response with a $0.5\ \mu\text{S}$ amplitude.

When using SCRs, one must consider that the EDA signal is subject to habituation: as reported in (Dawson et al., 2007), SCRs may not be observed after $2\text{--}8$ stimulus presentations, and the SCR amplitude can decrease by $0.01\text{--}0.5\ \mu\text{S}$ per stimulus presentation.

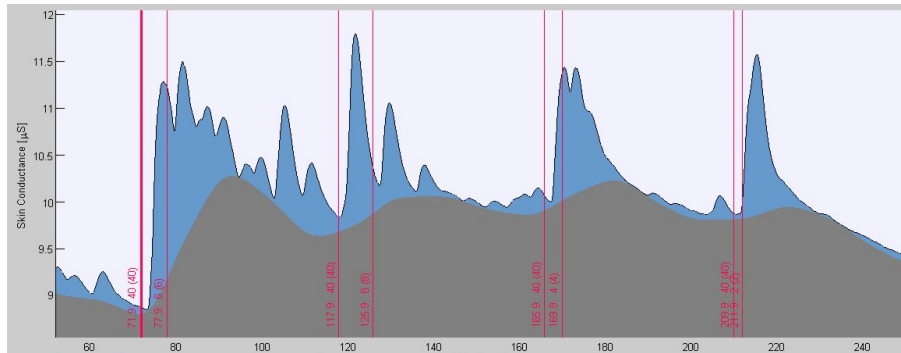


Figure 1.4: Screenshot from Ledalab, a Matlab-based software to decompose the EDA signal into SCR and SCL based on (Benedek and Kaernbach, 2010). The tonic (gray) and phasic (blue) components of the EDA signal are shown. The red vertical lines represent the discrete stimuli.

1.3.2 Facial muscle activity and eye-blink startle response

Muscle activity is measured through *electromyography* (EMG), which is a signal representing the electrical discharges of contracting muscle fibers. The electrodes employed for facial muscle activity detection in psychophysiology are able to detect subtle, subconscious muscle contractions that can go unnoticed by an experimenter observing the user's face, and that are strongly related to positive or negative emotional valence (Andreassi, 2007). Muscles like the *zygomaticus major* (located in the cheek), the *corrugator supercilii* (located in the forehead) and the *orbicularis oculi* (located around the eye) help to distinguish between positively- and negatively-valenced high arousal emotions (related to excitement and stress respectively). Rani et al. (2007) used the activity of the *masseter* muscle (located in the jaw) as indicator of anxiety, reporting various results in the literature. The suggested electrode placement for facial EMG are shown in Figure 1.5. Muscles from other parts of the human body, such as the *trapezius*, has also been used in the literature as indicator of emotional stress (Healey and Picard, 2005).

The analysis of facial muscle activity can also be useful to assess other physiological responses relative to anxiety and stress, such as the *startle* response. It involves a quick closing of the eyes, accompanied by contraction of various muscles, stiffening of the head, dorsal neck, body walls and limbs, as if to protect from a predator (Graziano and Cooke, 2006). Psychophysiology research mostly focused on the *eye-blink startle response* (EBSR), which is the fastest and more stable component of the startle response (Landis and Hunt, 1939). It is typically measured as the magnitude of the EMG recorded from the orbicularis oculi region, beneath the left or right eye (Blumenthal et al., 2005). EBSR has been successfully used in psychophysiology and biological psychiatry to assess the intensity of negative emotions such as contextual fear, and contextual and anticipatory anxiety (e.g., Grillon et al., 1991; Grillon, 2002; Baas et al., 2004), or more generally emotional valence (e.g., Baas et al., 2004; Cornwell et al., 2006; Mühlberger et al., 2008). In particular, Lang (1995) hypothesizes an inverse relationship between the intensity of EBSRs and the valence of a person's emotional state. Other features that are employed in the literature are blink latency (e.g., Bradley et al., 1990; Grillon et al., 1991) and frequency (e.g., Fairclough and Goodwin, 2007). Startle response is commonly elicited in laboratory studies through short bursts of intense white noise at 100–110 dB (e.g., Delgado et al., 2009; Globisch et al., 1999) or light flashes (e.g., Bradley et al., 1990). As reported by Globisch et al. (1999), researchers also used slides, odors and sounds to replicate startle responses.

Muscle activity is measured (in *microVolt*, μV) by placing electrodes on the skin above the considered muscles (*surface EMG*). The small contractions of the facial muscles, which are evaluated for emotion detection, generate EMG values that are typically in the 1–50 μV range. Since most of the activity of interest is between 20 and 200 Hz (Andreassi, 2007), a band-pass filter must be applied to the raw signal. In addition, a narrow band-stop filter centered around 50 or 60 Hz needs to be applied to remove electrical noise coming from the power line. Then, the filtered signal has to be rectified, typically employing *root mean square* (RMS) calculations. From the rectified signal, the following features can be extracted:

- Mean EMG value, calculated over a given time interval. It is employed to measure the mean activity

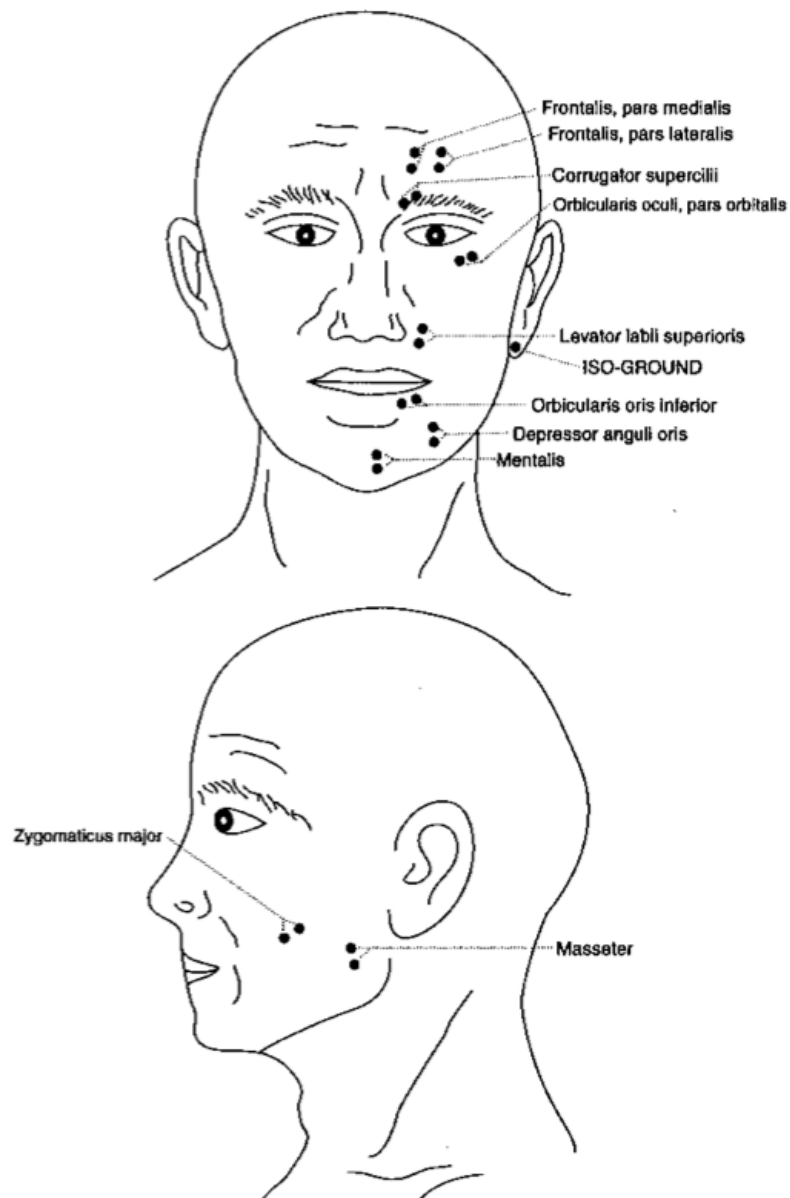


Figure 1.5: Suggested electrode placements for facial EMG measurements, from (Tassinari et al., 2007).

of facial muscles during the experimental task;

- Mean EMG amplitude/magnitude: the mean value of the differences between the EMG values at response onsets and response peaks measured in the 21–120 ms time window after each startle probe (Blumenthal et al., 2005). The difference between amplitude and magnitude is the same already discussed in Section 1.3.1.

To obtain a considerable economy in terms of data acquisition and signal processing, a smoothing algorithm (e.g., a simple moving window) can be applied (Tassinari et al., 2007) when calculating the mean EMG value over a time interval. However, when the goal is measuring the startle response, the latency that the smoothing algorithm can introduce must be taken in consideration. Figure 1.6 presents various representations of the EMG signals.

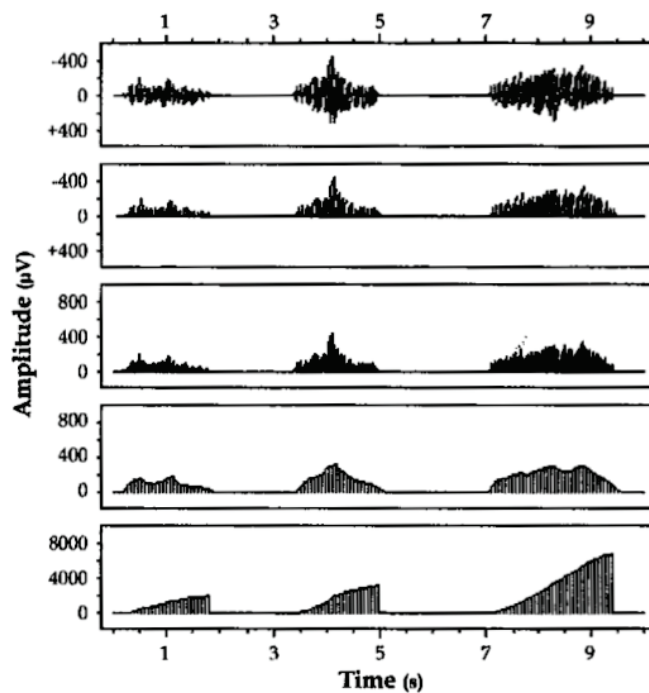


Figure 1.6: Various representations of the EMG signals, from (Tassinari et al., 2007). From top to bottom, (i) raw EMG signal; (ii) and (iii) rectified signal; (iv) smoothed EMG; (v) integrated EMG.

1.3.3 Cardiovascular system activity

The activity of the cardiovascular system can be measured through various instruments. In the literature, *blood volume pulse* (BVP) and *electrocardiography* (ECG) are the most employed. BVP is related to the amount of blood flowing into the peripheral vessels, like the ones in the fingers, and is generally measured through a photoplethysmograph (Figure 1.7). It generates a subtle IR light that is reflected by the capillaries under the finger's skin by an amount that is proportional to the quantity of blood flowing in the capillaries. The ECG signal represents the electrical activity of the heart, and is measured through electrodes placed on the user's chest. The main disadvantage of measuring HR through a PPG is that the sensor is very sensitive to hand motions, and encumbers the movement of the user's hand. ECG electrodes are more robust against motion artifacts, but they must be placed in direct contact with the skin of the chest, which is less practical and may also be uncomfortable for some users. Raw BVP and ECG signals are shown in Figure 1.8.

Heart rate (HR), measured in *beats per minute* (bpm), can be easily calculated from the BVP or the ECG signals by counting the number of peaks per minute, and is one of the most common physiological parameters that can be employed in emotion detection. Increases in HR are generally related to emotional



Figure 1.7: A commercial photoplethysmograph.

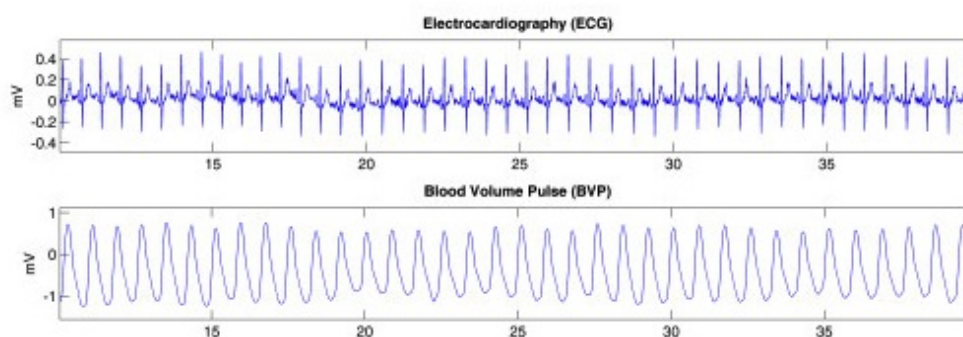


Figure 1.8: ECG and BVP signals.

activation, and have been used in the literature as a correlate of arousal (Mandryk and Atkins, 2007). As reported by Boucsein (2006), compared to EDA, HR is well suited as an indicator for the higher arousal range and for pronounced and often somatically determined arousal processes.

Other features relative to HR have been explored in the literature as more precise measures of certain characteristics of ANS activity, such as *heart rate variability* (HRV). It describes the natural variability of both instantaneous heart rate and *interbeat intervals* (IBIs) (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996). As reported by the studies cited by Muth et al. (2012), HRV is regulated by ANS activity, and it is one of the most often studied physiological indexes of mental workload: a higher HRV is positively related to sympathetic–parasympathetic balance, which is a strong index of effective stress management (Brouwer et al., 2011). The acquisition of HRV data is now possible in most environments, using both BVP and ECG.

In general, HRV analysis takes in consideration the frequency power of the low frequency (0.01–0.08 Hz) and high frequency (0.15–0.5 Hz) bands. The low frequency band is primarily considered a measure of sympathetic activity (Akselrod et al., 1981), while the high frequency band is dominated by *respiratory sinus arrhythmia* (RSA), which describes the phenomenon of cyclic HR changes in phase with respiration: HR increases during inhalation and decreases during exhalation (Delgado et al., 2009, 2010). RSA reflects alteration in the vagal function, and has been validated as an index of PNS activity when respiratory influences are accounted for (Grossman and Taylor, 2007). Jönsson (2007) reported that high state anxious participants had higher RSA magnitude than low state anxious individuals. Increasing RSA magnitude was also observed when approaching a stressful event (Fuller, 1992).

Slow and irregular variations in HR, such as *cardiac defense response*, are also studied in the literature. Cardiac defense response refers to heart activity in response to intense or aversive stimulation, and consists of a complex response pattern, observed within 80 s after stimulus onset: (i) a short latency acceleration and subsequent deceleration (with a peak around 3 s after the stimulus), and (ii) a long latency acceleration and subsequent deceleration (with a peak around 30 s to 40 s after the stimulus) (Vila et al., 2007). As

observed in the studies reported by Delgado et al. (2009), cardiac defense response has components that are differentially mediated by sympathetic and parasympathetic mechanisms: the short latency component is controlled by vagal influences and the dominance of PNS activity, whereas the long latency component is controlled by reciprocal SNS and PNS influences, with a dominance of SNS activity.

Another physiological parameter that can be calculated from the raw BVP signal is the *BVP amplitude* (BVPA), which is the distance between local maximum and minimum of the signal. Since the size of the blood vessels is controlled by both SNS and PNS, BVPA measurements can be employed as index of sympathetic arousal. An increase in the BVPA indicates decreased sympathetic arousal and greater blood flow, and thus higher temperature at the fingertips.

In general, all commercial BVP and ECG sensors can automatically provide HR values in real time. Alternatively, in the literature several algorithms for HR and BVPA calculation are proposed. For example, Zong et al. (2003) proposed an open-source algorithm which obtain HR values by detecting peaks in the BVP signals, which can also be used for BVPA calculation. To proceed with the analysis of HRV, a *fast Fourier transformation* (FFT) must be applied to translate the HR data from the time domain to the frequency domain. This approach can lead to tens of seconds of latency, and one must consider this limitation when using HRV analysis for real-time emotion detection. Furthermore, as reported in (Hoover and Muth, 2004), there is currently no known method for automated IBI artifact correction and detection. As Hoover and Muth observed when describing their RSA-based arousal meter system, even a single IBI error can cause the system to report incorrect arousal levels for a short period surrounding the wrong IBI value.

1.3.4 Respiratory system activity

Respiration, like peripheral temperature, is strongly related to cardiovascular system activity, as already observed when discussing RSA. In the literature, the *respiratory rate* (RR, also called *respiratory frequency*, f) and *respiratory amplitude* (RA) are employed as measures of SNS activity. As reported in (Katsis et al., 2008), emotional excitement and physical activity lead to faster and deeper respiration, whereas peaceful rest and relaxation lead to slower and shallower respiration. A state of stress would therefore be indicated by frequent respiration; however, sudden stressors, such as startle, tend to cause momentary cessation of respiration. As reported in (Cacioppo et al., 2007), RR (measured in *cycles per minute*, cpm) is among the least sensitive metrics for respiratory data analysis, so many researchers proposed, since the beginning of the 20th century, other time-domain measurements relative to respiration, such as *inhalation time* (T_i), *exhalation time* (T_e), and *duty cycle*, i.e., the ratio between the inspiration time and the sum of inspiration and expiration time. Within subjects standard deviations (SDs) of these measurements, as well as other measurements of their variability, have been frequently used to characterize respiratory variability (Benchetrit, 2000; Van Den Wittenboer et al., 2003), which seems to be negatively correlated to anxiety (Van Den Wittenboer et al., 2003; Van Diest et al., 2006).

The respiratory signal can be recorded using an elastic band placed on the person's girth (Figure 1.9). To measure RR, FFT can be applied to the raw respiratory signal; the peak frequency can then be taken as the mean RR value for the time window considered by each FFT application. This approach removes the effects of the small artifacts caused by brief inspirations and expirations on the respiratory signal. As reported in Section 1.3.3, the FFT leads to higher latency than a peak detection algorithm, which can also be employed for the various time-domain measurements previously described.



Figure 1.9: On the left, a respiratory sensor; on the right, an example of respiratory signal.

1.3.5 Brain activity

The electrical activity of the brain is measured through *electroencephalography* (EEG), i.e., the sum of various electrical signals (called *waves*) that can be identified by analyzing the EEG spectrum (Andreassi, 2007). The most relevant to affective computing are the following:

- α waves (8–13 Hz): are observed during relaxation, especially with closed eyes;
- β waves (14–30 Hz): are observed during mental or physical workload;
- γ waves (around 40 Hz): are observed as a rhythmic activity that occurs in response to sensory stimuli (auditory clicks or light flashes);
- θ waves (4–7 Hz): are observed during states of displeasure, pleasure and drowsiness in young adults.

Event-related brain potentials (ERPs) are a common measure of discrete brain activations derived from EEG recordings. Unlike EEG, which represents spontaneous brain activity, ERPs are generated as a response to specific stimuli, and are calculated over an average of a number of samples (Andreassi, 2007). To be more precise, the EEG signal is recorded over epochs time-locked to event repetitions, then the digital EEG values recorded during epochs are averaged to remove signal noise. This signal filtering procedure, while very effective, limits the use of ERPs during real-time emotion detection where stimuli replication cannot be controlled; furthermore, ERPs waveforms and latency is variable among people. However, ERP can be successfully employed in stress and anxiety recognition: anxious people have shown different ERP patterns with respect to non-anxious people during an emotion-word Stroop task (Sass et al., 2010). One of the most used ERP is the *P300*, or *P3* (it refers to the onset latency from the stimulus, which in this case is around 300 ms): the waveform is easily detectable, and it can be observed in nearly all people.

Another EEG feature employed in the psychophysiology literature to measure stress and mental workload is the frontal *alpha asymmetry* (AA), which is the ratio of α waves activity between the frontal lobes of the left and right hemispheres. Unlike ERPs, AA can be employed in real-time emotion detection, because no stimuli repetition is required. According to the studies reported in (Brouwer et al., 2011), differences between frontal brain activity in left and right hemisphere are correlated with stress. Furthermore, the direction of the asymmetry may depend on the individual's coping strategy (Brouwer et al., 2011): a greater left or right alpha power seems to be related to an avoidance-driven or approach-driven strategy respectively.

EEG is measured by placing electrodes (from 2 up to as many as 256) over a person's head. Electrode position are described in the *International 10–20 System* proposed by Jasper (1958) (Figure 1.10). The name refers to the fact that electrodes are placed at sites 10% and 20% from four predefined points on the scalp. As reported in (Pizzagalli, 2007), in recent years the 10–20 system has been extended by adding more and more electrode positions. In general, other than the raw EEG signal, many commercial EEG sensors automatically provide α wave values in real time; alternatively, a band-pass filter (in the 8–13 Hz range) can be applied to the EEG signal. To calculate the α power values, the two signals must be translated into the frequency domain using a FFT. Finally, the AA is obtained by performing a subtraction between the two log-transformed α power values (Brouwer et al., 2011). The absolute value of AA can also be considered, since the stress coping strategy varies among people.

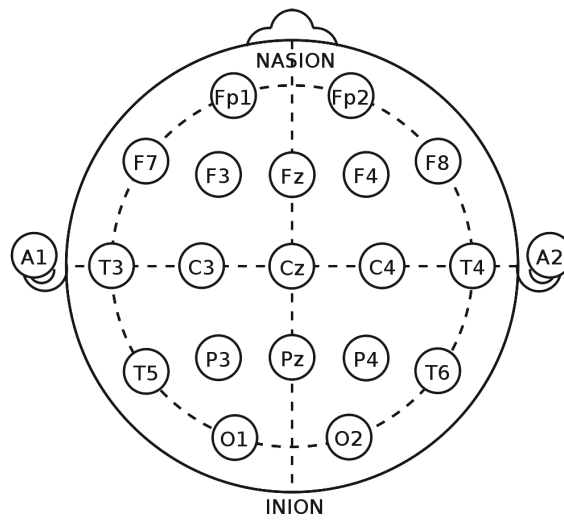


Figure 1.10: Electrodes position and labels in the International 10–20 System.

2

Affective State Detection and Its Applications to Stress and Anxiety Management

In this chapter, we discuss affective state detection, with particular attention to stress and anxiety. In Section 2.1, we describe the most common algorithms and techniques employed in the literature to interpret physiological signals as indices of intensity of emotions. In particular, we describe the basic concepts behind each technique with examples of its use in the affective computing literature. Then, in Section 2.2, we examine the emotion detection systems and affective computing systems proposed in the literature, and critically discuss what physiological signals and what algorithms and techniques are employed, in relation with the accuracy afforded by each system. Finally, in Section 2.3, we discuss the possible applications of affective computing systems for stress and anxiety management, also referring to the systems proposed in the literature and examined in Section 2.2.

2.1 Algorithms and Techniques for Automatic Stress Detection

Since our goal is to gather information about users' affective state from their physiological responses, algorithms and methods are needed to transform physiological data into quantities representing the intensity of emotions, in particular data fusion and pattern recognition techniques. Data fusion algorithms combine data from multiple sources to achieve inferences. In our case, various physiological data can be employed to calculate one or more indices, which can later be labeled using pattern recognition algorithms to associate them to single or multiple emotion values. For example, EDA and cardiovascular activity can be combined to generate a single index that is later interpreted as representing a low, medium or high level of arousal. The following sections will introduce the most common techniques and algorithm employed in the literature to gather information about users' emotions from physiology.

2.1.1 Fuzzy logic

Fuzzy logic (FL) (Zadeh, 1965) differs from Boolean algebra in the value of truth that can be associated to variables. Boolean logic variables can assume only two discrete values (true or false, 0 or 1), while FL variables can have a continuous degree of truth (which, albeit similar, is conceptually different from probability) that can vary between 0 and 1, which represent the degree of membership to a fuzzy set. The association of a fuzzy element to a fuzzy set is defined by a *membership function*. FL can be used to describe mathematically an approximate reasoning that are better associated to linguistic rather than numeric variables. Furthermore, it can easily represent continuous processes, like changes in the affective state, that are not easily broken into discrete segments, and in which the change of state from one linguistically-defined level to the next is not clear (Cox, 1992). For example, Mandryk and Atkins (2007) define three fuzzy sets to determine if HR values (normalized between 0 and 100) can be considered low, medium, or high.

Once input signals are fuzzified, inference systems like the Mamdani system (1975) employ fuzzy rules (simple conditional statements, e.g., “if EDA is high and HR is high then arousal is high”) to map them to a fuzzified output associated to one or more fuzzy sets (“high arousal”, “medium arousal”, “low arousal”). Mandryk and Atkins employed two different Mamdani system, the first one to map physiological signals to the arousal–valence space, and the second one to map arousal and valence fuzzy values to five emotion values.

2.1.2 Bayesian networks and influence diagrams

Bayesian networks (BNs) were first described in (Pearl, 1985). The name comes from Bayes’ theorem, which describes the relations between two conditional probabilities $P(A|B)$ and $P(B|A)$, where A and B are two events which can take place respectively with probability $P(A)$ and $P(B)$. BNs represent a set of random variables and their conditional dependencies via a direct acyclic graph. For each node, representing a random variable, a probability function is given. From a BN, two types of inference can be done: predictive support through top–down reasoning, and diagnostic support through bottom–up reasoning. For example, a node representing the probability of “presence of stress” can derive its value from the combined presence of high SCR magnitude and high masseter activity. In an ACS, predictive support would likely be the most used type of inference. Nakasone et al. (2005), for example, derive the probability that a person is feeling frustrated by the probability of high arousal and negative valence, which are respectively determined by EDA, EMG and the status of the application used.

Influence diagrams (IDs) are a generalization of BNs that can represent and solve decision problems. They introduce decision nodes that are activated on the base of on parents’ outcome, and utility nodes, a quantification of benefits and penalties, which are a function of parents’ outcome. IDs can be used in ACSs to model the activation of user assistance functionalities (Liao et al., 2006).

2.1.3 Hidden Markov models

Hidden Markov models (HMMs) are the simplest dynamic Bayesian networks, and are especially useful in temporal pattern recognition (e.g., in the case of a real–time stream of input data). A HMM consists of a set of hidden internal states and a set of possible observations. Internal states are linked by state transition probabilities; internal states and observations are connected by output probabilities. A common goal is to generate a HMM structure (representing all the different states of an application, e.g., high– and low–stress states) from a set of training data (physiological data related to a known stress state), which can be then used to analyze in real–time the input data (the physiological data of the trainee) and estimate the stress level of the user. For example, Scheirer et al. (2002) employed a HMM to classify physiological signals for frustration detection.

2.1.4 Naïve Bayes classifiers

A *Naïve Bayes classifier* (NBC) assumes that all the features of a sample (e.g., the physiological measurements) are independent from each other. This method computes the posterior probability of each class (e.g., “high stress”, “medium stress” and “low stress”) given the values of attributes of a test instance; the classifier predicts that the instance belongs to the class with the highest posterior probability. The conditional probability is estimated by the probability mass function using training physiological data. Zhai and Barreto (2006) tested a NBC, which employed various physiological measurements like HR, BVPA, and EDA to detect if a person is stressed or relaxed.

2.1.5 Support vector machines and linear discriminant analysis

Support vector machines (SVMs) are employed to subdivide the input data into two or more classes by using one or more hyperplanes defined through a training session. Typically, a SVM generates the hyperplanes that maximize the distance from the nearest training points of any class. For example, Katsis et al. (2011) developed and evaluated a SVM that subdivides a set of multi–dimensional points described by 8

physiological measurements derived from BVP, EDA and respiratory data in five sets (“relaxed”, “neutral”, “startled”, “apprehensive”, and “very apprehensive”).

Similar to SVMs, *linear discriminant analysis* (LDA) finds a linear transformation (the discriminant function) that maximizes class separation: in the case of 2D data, it describes the function of the line on which the data points are projected, and which maximize the minimal distance between the projections belonging to the different classes. For example, Healey and Picard (2005) extracted 22 features from EMG, EDA, respiratory, and ECG data, and developed a LDA to classify multi-dimensional data values into three classes representing low, medium, and high stress levels.

2.1.6 K-nearest neighbor algorithm

The *k-nearest neighbor* (KNN) algorithm employs a training data set, already partitioned in classes, and the input data set. Given an instance of input data, it finds the k nearest training data instances; the class with the higher number of instances among the k is associated to the input data instance. For example, in (Lisetti and Nasoz, 2004) the k training instances are k vectors collecting EDA, HR and skin temperature data, divided into six classes, representing six basic emotions.

2.1.7 Decision tree classifiers

The idea behind *decision tree classifiers* (DTCs) is to break up a complex decision into a union of several simple ones. A data set is split, with each part analyzed at each level of the tree from the root to the leaf representing the output value. A DTC is employed, for example, by Rani et al. (2007) for their anxiety detection system. The nodes in the DTC test for the different values derived from BVP, peripheral temperature and EDA data, and the leaves represent each one a different anxiety level.

2.2 Automatic Emotion Detection and Affective Computing Systems

In the affective computing literature, various biosensor-based emotion detection systems as well as ACSs have been proposed; some of the most recent are reported in Table 2.1. In many cases, their purpose is not limited to stress and anxiety detection: many of them are developed to recognize a wide variety of emotions, such as joy and surprise. In the following, we discuss the most recent ACSs and emotion detection systems proposed in the literature, highlighting the limitations emerged from our literature review.

Looking at Table 2.1, one can immediately notice how emotion detection accuracy is generally not very high: only the systems proposed by Healey and Picard (2005), Zhai and Barreto (2006), Rani et al. (2007), Van Den Broek (2011), and Fleureau et al. (2012) reached or surpassed a detection accuracy threshold of 90%. However, it must be considered that both the studies by Healey and Picard, and by Rani et al. involved only one and five participants respectively. The emotion detection techniques employed seem not to have an impact on these results, since each system is based on different techniques (LDA, SVM, DTC), which are also used by systems with lower emotion detection accuracy. Each technique, however, has its strengths and weaknesses. Rani et al. (2007) reported differences in error percentage when comparing the accuracy of DTCs and FL. The former obtained very good results when the training data is optimal; the latter, while reporting higher error percentages, was more robust than the DTC algorithm in case of bad training sets. Another algorithm comparison is performed by Zhai and Barreto (2006), who evaluated the stress detection accuracy of a NBC, a DTC and a SVM in a Stroop task. In their study, the SVM came out as the most accurate technique, and the authors explained the result with the greater ability of SVMs in dealing with complex pattern distributions (which is the case when dealing with multiple physiological signals). Katsis et al. (2008) obtained similar results when comparing the accuracy of a SVM and a FL system trained with a neural network for driving tasks. Authors observed that the SVM obtained a slightly higher accuracy (79% vs. 76%) when detecting high and low stress levels, disappointment and euphoria. Hernandez et al. gave particular attention to the impact that SVM customizations (e.g., automatic selection of training samples) can have on emotion detection accuracy. Lisetti and Nasoz (2004) compared LDA, KNN and a backpropagation algorithm (an algorithm based on the use of neural networks) when detecting

Proposal	Emotion(s)	Physiological signal(s)	Technique(s)	Results
Oshuga et al. (2001)	Phasic work stress	ECG, BVP, facial temp.	No technique	Significant differences in phys. data between high-tension and monotonous stress
Scheirer et al. (2002)	Frustration	BVP, EDA	HMM	Accuracy: 67.40%
Kim et al. (2004)	Sadness, anger, stress	ECG, EDA, skin temp.	SVM	Accuracy: 78.43%
	Sadness, anger, stress, surprise			Accuracy: 61.76%
Lisetti and Nasoz (2004)	Sadness, amusement, fear, surprise, frustration, amusement	BVP, EDA, skin temp.	KNN	Accuracy: 72.30%
			LDA	Accuracy: 75.00%
			Backprop. alg.	Accuracy: 84.10%
Healey and Picard (2005)	Stress	ECG, EDA, EMG, resp.	LDA	Accuracy: 97.40%
Liao et al. (2005)	Stress	BVP, EDA, skin temp.	ID	Qualitative assessment of provided assistance
Nakasone et al. (2005)	Fear, frustration, relaxation, joy, excitement	EDA, EMG	BN	N.D.
Slater et al. (2006)	Social stress	ECG, EDA	No technique	Significant differences betw. socially anxious and confident
Zhai and Barreto (2006)	Stress	BVP, EDA, skin temp., pupil diam.	NBC	Accuracy: 78.75%
			DTC	Accuracy: 88.02%
			SVM	Accuracy: 90.10%
Friedman et al. (2007)	Stress (arousal), relaxation	EDA	No technique	No positive/negative loop observed
Jönsson (2007)	State anxiety	ECG	No technique	Significant differences betw. high/low state anxious
Mandryk and Atkins (2007)	Fun, challenge, boredom, frustration, excitement	ECG, EDA, EMG	FL	Significant differences among playing against a friend/stranger/computer
Rani et al. (2007)	Anxiety	ECG, BVP, EDA, EMG, skin temp.	FL	Accuracy: 94.64%–56.60%
			DTC	Accuracy: 78.48%–74.88%
Katsis et al. (2008)	High/low stress, euphoria, disappointment	EMG, ECG, resp., EDA	SVM	Accuracy: 79.3%
			FL + neural network	Accuracy: 76.7%
Chanel et al. (2011)	Engagement, anxiety, boredom	EDA, BVP, EEG, skin temp., resp.	LDA, Quadratic DA, SVM, & feat.-selection algorithms	Accuracy: 48%–59%, 63% with data fusion
Hernandez et al. (2011)	Stress	EDA	Modified SVM	Accuracy: 73.41% with SVM corrections
Van Den Broek (2011)	Arousal–valence space	ECG	No technique	Explained variance: 90%
	Joy, anger, relaxation, sadness, neutral (high arousal), neutral (low arousal)			Explained variance: 18%
Bouchard et al. (2012)	Arousal	EDA, HR	Function for data fusion	N.D.
Fleureau et al. (2012)	Valence	EMG, EDA, BVP	Gaussian Process Classifiers	Prediction of binary valence: 80%–90%
Muth et al. (2012)	Cognitive workload	ECG	No technique	Correlation betw. RSA residuals & NASA-TLX

Table 2.1: Some of the ACSs presented in the literature since 2001 that use physiological sensors to perform emotion detection.

six basic emotions (sadness, anger, fear, surprise, frustration, and amusement) during video watching. They observed that no algorithm was better than the other two for all of the six emotions.

When data fusion and classification are employed, the accuracy of emotion detection is generally assessed through confusion matrices (Kim et al., 2004; Healey and Picard, 2005; Zhai and Barreto, 2006; Chanel et al., 2011), mean error (Rani et al., 2007) or mean successful recognition (Katsis et al., 2008; Lisetti and Nasoz, 2004; Scheirer et al., 2002) rates. Some of the ACSs reported in Table 2.1 did not employ data fusion and classification algorithms at all, but focused on directly mapping physiological functions to affective states (e.g., Van Den Broek, 2011; Friedman et al., 2007; Jönsson, 2007; Liao et al., 2006; Muth et al., 2012; Oshuga et al., 2001; Slater et al., 2006). In these studies, the goal is to verify if the considered physiological signals were able to distinguish between tasks of different stress level (Oshuga et al., 2001) and between more or less anxious participants (Friedman et al., 2007; Jönsson, 2007), or if the physiological signals correlate with subjective stress assessment (Liao et al., 2006; Muth et al., 2012). A few studies evaluated the effectiveness of the proposed system by employing a control condition, in which the system is simply not used (i.e., emotion detection is not performed). For example, in (Bouchard et al., 2012) authors evaluated the effects of an ACS-enhanced military training game by comparing it to the original version of the same game. To the best of our knowledge, no studies have employed a placebo condition, i.e., a control condition in which users' emotions, unbeknown to them, are determined pseudo-randomly instead of taking into account their physiological sensor readings.

In the considered ACSs, EDA is the most employed physiological signal to detect stress and anxiety (e.g., Healey and Picard, 2005; Scheirer et al., 2002; Slater et al., 2006). This is not surprising, because EDA is strongly related to SNS activity (as reported in Section 1.3.1), and also is very robust against artifacts, making it a signal that is easy to process. The activity of cardiovascular and respiratory systems is another physiological function widely used in the considered ACSs and, as reported in Sections 1.3.3 and 1.3.4, in the affective computing literature. BVP sensors are often employed because, like the EDA ones, they are less invasive than other sensors (such as the ones employed for measuring EMG and EEG) since they can be applied to users' hands instead of other sensible body parts like users' face. In fact, in many studies (e.g., Zhai and Barreto, 2006) authors highlighted the fact that they have chosen physiological sensors that are as less invasive as possible. Signals from BVP sensors, however, can be prone to motion artifacts unlike ECG electrodes which, albeit more bothersome to wear, can be also employed when users are required to move. While invasive, EEG seems to be an important physiological signal for short-term emotion assessment (Chanel et al., 2011). Other physiology signals, such as trapezius and masseter muscles activity (Rani et al., 2004; Healey and Picard, 2005), are less employed for stress detection. *Pupil diameter* is another physiological function rarely employed in the considered studies. Zhai and Barreto (2006), however, reported that the inclusion of pupil diameter had a strong positive effect on the detection accuracy of their system, suggesting that the inclusion of physiological signals rarely employed in the affective computing literature is worth considering to improve the effectiveness of ACSs and emotion detection systems.

Many ACSs reported in Table 2.1 (e.g., Chanel et al., 2011; Fleureau et al., 2012; Zhai and Barreto, 2006) employ multiple physiological signals to improve the accuracy of emotion detection. Chanel et al. (2011) performed a fusion of EEG features and peripheral signals (EDA, HR, HRV, skin temperature and respiratory data), and successfully increased the accuracy of emotion detection in an adaptive *Tetris* game with respect to the accuracy obtained when considering EEG or peripheral signals alone. Zhai and Barreto (2006), after obtaining good accuracy results with SVMs and BVPA, EDA, skin temperature and pupil diameter, re-tested their ACS while removing each time one of the physiological signals. They observed that the detection accuracy always decreased down to 61% when removing pupil diameter measurements. When removing other physiological signals such as EDA and BVP, however, SVMs accuracy only decreased by less than 1%. This is somewhat surprising: in fact, Healey and Picard (2005) observed that EDA is one of the strongest correlates of stress during driving tasks. On the other hand, some of the systems in Table 2.1 focus on a very limited array of emotions and physiological signals. For example, Jönsson (2007) and Muth et al. (2012) employed RSA to evaluate state anxiety and mental workload respectively. Other studies (e.g., Friedman et al., 2007; Hernandez et al., 2011) focused instead only on EDA to detect arousal and stress. From this literature review, it is clear that further investigation is needed about possible differences in emotion detection accuracy among ACSs that employ a different number of sensors, e.g., an EDA-based system and a system based on EDA, BVP, and EMG sensors.

Nakasone et al. (2005), unlike other systems in the literature previously described, employed the state of a simple virtual card game (ranging from very favorable to very unfavorable to the player) together with player's physiology to detect his/her emotions. The use of the game state allowed authors to more easily hypothesize the player's positive or negative appraisal of the current situation of the game.

2.3 Applications of Affective Computing to Stress and Anxiety Management

Various studies in the literature observed correlations between stress, performance and human error (e.g., Hancock and Szalma, 2008). One of the goals that HCI focuses on is the reduction of users' stress and frustration levels when using an electronic device, may it be a desktop computer, a mobile phone or the controls in a modern airplane cockpit. For example, a state of relaxation may help increasing the productivity of a user interacting with a word processor on a PC. To this purpose, easy-to-use graphic interfaces, multimodal feedback, and interactive helpers have been introduced over the years. Other contexts may require reaching and maintaining a certain stress level: for example, in a driving task, maintaining an optimal level of stress may favor concentration, reducing errors caused by distraction. In all these cases, an accurate recognition of stress and anxiety is crucial. In the following, various examples of biosensor-based affective computing applications are presented.

2.3.1 Training and biofeedback

Training allows the acquisition of knowledge, competence and skills through direct or indirect experience. During mass emergency training, for example, first responders could learn how to provide medical care by being immersed in real-world simulations that accurately replicate, possibly on a large scale, a dangerous environment that poses various risks. As another example, training activity can help trainees to develop coping skills to reduce anxiety and to maintain an optimal level of performance under stress. This particular training activity is called *stress inoculation training* (SIT) (Meichenbaum and Deffenbacher, 1988).

Training systems can exploit immersive VEs instead of real-world drills. A VE is a computer-generated, generally three-dimensional representation of a setting, such as 3D virtual exhibitions, virtual home/interior design, and video game levels. Immersive VE-based training systems allow one to perform the training activity with less expense and with no risks. An ACS could be applied to a VE-based training application for real-time detection of users' stress level, caused by the events simulated during the training activity. The ACS could then modulate the task difficulty and the amount of information given to users on the basis of their emotional state, to help them to maintain an optimal level of stress. Affective computing is thus a perfect instrument to be applied to SIT (Popović et al., 2009). Furthermore, the use of an ACS allows for the definition of a physiological profile unique for each single user and tailor the training activity to each person.

Three professional biofeedback organizations, the Association for Applied Psychophysiology and Biofeedback (AAPB), the Biofeedback Certification International Alliance (BCIA), and the International Society for Neurofeedback and Research (ISNR), define biofeedback as the following:

Biofeedback is a process that enables an individual to learn how to change physiological activity for the purposes of improving health and performance. Precise instruments measure physiological activity such as brainwaves, heart function, breathing, muscle activity, and skin temperature. These instruments rapidly and accurately "feed back" information to the user. The presentation of this information—often in conjunction with changes in thinking, emotions, and behavior—supports desired physiological changes. Over time, these changes can endure without continued use of an instrument. (The Association for Applied Psychophysiology and Biofeedback, 2011)

The use of biofeedback in relaxation training and stress treatment is widely reported in the literature. The goal of this biofeedback activity is to improve trainees' health and wellness by increasing their ability to relax, and to make them learn how to better cope with stress. Indeed, in some cases high levels of arousal are

hallmarks of anxiety disorders and stress (Chandler et al., 2001), and pose a threat to well-being, safety and security (Carr, 2006). Interventions such as biofeedback-assisted relaxation training could reduce stress-related symptoms (Chandler et al., 2001), improve quality of life (Pistoia et al., 2013) and increase feelings of well-being (Chandler et al., 2001; Critchley et al., 2001; Pistoia et al., 2013). In the literature, various examples of biofeedback, integrated for example into SIT activities, have been proposed (e.g., Gorini et al., 2010; Bouchard et al., 2012) (Figure 2.1).



Figure 2.1: A screenshot from the biofeedback game presented in (Gorini et al., 2010). A reduction in heart rate reduces fire intensity, and vice versa.

Biofeedback-based relaxation training is also used in the treatment of medical conditions, e.g., chronic pain (e.g., Jensen et al., 2009), fibromyalgia (e.g., Buckelew et al., 1998), migraine (e.g., Fentress et al., 1986; Pistoia et al., 2013; Vasudeva et al., 2003), and hypertension (e.g., Hunyor et al., 1997). In these cases, biofeedback leads to positive effects on patients' health and wellness that, in some cases, are at least comparable to traditional therapies (Buckelew et al., 1998; Jensen et al., 2009; Vasudeva et al., 2003). To further improve the efficacy of the intervention, the integration of biofeedback with traditional therapies is often taken into consideration (e.g., Pistoia et al., 2013). The fact that biofeedback-based treatments, while effective, do not provide in some cases significant advantages over classic therapies led some authors (e.g., Fentress et al., 1986) to question whether the expense of biofeedback equipment is justifiable when simpler and less costly alternatives. For example, *relaxation response* (Benson and Klipper, 1975), i.e., a set of co-ordinated physiological changes related to a state of increasing relaxation, can be brought forth when a person focuses attention on a repetitive mental activity (e.g., repeating a word or a phrase) and passively ignores distracting thoughts (Fentress et al., 1986).

Jensen et al. (2009) report a common limitation of many biofeedback-based therapy studies in the literature, which often focus only on the short-term effects of the therapy on patients' medical conditions, without paying much attention to longer-term effects. This issue is also apparent in studies that apply biofeedback in the treatment of gastrointestinal and pelvic floor disorders. The reviews by Enck (1993) and Bassotti and Whitehead (1997) report how the application of biofeedback seems to improve, in the short term, numerous functional disorders of the gastrointestinal tract, especially those related to the lower part of the gut (Bassotti and Whitehead, 1997), but more research on the long-term effects of these treatments must be performed (Enck, 1993).

Attention deficit-hyperactivity disorder (ADHD) is another common target for biofeedback-based treatments; see the original proposal by Lubar and Shouse (1976), the review by Lubar Lubar (1991) and the meta-analysis by Arns et al. (2009). In particular, Lubar (1991) shows that biofeedback training, although time consuming, can lead to significant improvements. Arns et al. (2009) confirm that biofeedback in the treatment of ADHD can be regarded as clinically meaningful, also highlighting that the clinical effects seem to be stable and might even improve further with time.

To derive physiological indices of arousal and stress, the biofeedback literature often exploits EMG

signals from facial muscles and EDA (e.g., Bouchard et al., 2012; Critchley et al., 2001, 2004; Jensen et al., 2009; Reynolds, 1984). Peripheral temperature and cardiac parameters such as HR and HRV are also employed (e.g., Bouchard et al., 2012; Shusterman and Barnea, 2005). In some cases, EMG is employed as a direct measure for the specific physiological phenomena under treatment: for example, in (Foster, 2004), EMG from the masseter muscle (located in the jaw) was employed in the biofeedback-based treatment of chronic nocturnal bruxism; in almost all studies of pelvic floor disorders cited above, the activity of pelvic muscles was measured to provide biofeedback. Other studies, especially those dealing with the treatment of attention deficit and hyperactivity disorders, employ EEG to infer the affective state of the patient.

2.3.2 Performance and health monitoring during work

An high level of stress could negatively impact a worker's performance (Hancock and Szalma, 2008) as well as her physical health (Taveira and Smith, 2006). The recognition of states of excessive anxiety and stress in workers is a typical task of human resource management. However, skilled and experienced managers and psychologists are required, because the symptoms of excessive stress may not be immediately noticeable.

ACSs can automatically detect workers' stress level (e.g., Hernandez et al., 2011), and, for example, be integrated into PCs in the workplace to dynamically manipulate applications to continuously maintain an optimal stress level in workers, for example by providing user assistance or decreasing the difficulty of the task (Liao et al., 2006). In (Healey and Picard, 2005), authors describe an ACS used to detect excessive stress during driving tasks. The proposed system could be integrated in future cars to manage noncritical in-vehicle information systems, and to measure how different road and traffic conditions affect drivers.

2.3.3 Exposure therapy

Exposure therapy is a technique in behavior therapy intended to treat anxiety disorders and involves progressive exposure to the feared object or context in order to inhibit fear and overcome anxiety (Myers and Davis, 2007). The exposure therapist identifies the cognitions, emotions and physiological arousal that accompany a fear-inducing stimulus, and attempts to break the pattern of escape that strengthens the fear response, through measured exposure to progressively stronger stimuli until habituation is reached (De Silva and Rachman, 1981). Exposure therapy can be successfully used not only in the treatment of generalized anxiety disorders, but also in treatment of post-traumatic stress disorders and phobias. However, exposure therapy, despite its demonstrated efficacy, remains underutilized (Jaeger et al., 2009), also because it is not completely accepted by psychologists, who seems to not understand it, or are uncomfortable with it (Becker et al., 2004).

The use of VEs paired with affective computing allows for an easier to implement, safer and more controlled therapy than, for example, *in vivo* exposure therapy (in which patients directly confront themselves with the feared object, activity, or situation) or imaginal exposure therapy (in which patients have to revisit traumatic memories related to their fear). In the literature, there are various examples of exposure therapy paired with ACSs for patients' affect detection. For example, Katsis et al. (2011) propose a novel architecture, called *INTREPID*, to be employed during exposure therapy sessions to detect affective states in patients with anxiety disorders. Bouchard et al. (2012) describe a biofeedback-based video game for stress management therapy in the military field. The video game is designed to elicit stress, and employs HR and EDA to provide feedback to users about their level of arousal. Authors show that the video game was able to effectively increase stress management skills in soldiers.

3

Physiology–Based Evaluation of Interaction Devices

In Chapter 1, we described some of the most common physiological signals employed in affective computing research. In this chapter, we employ some of these signals, with a particular emphasis on EMG, to measure users' bodily responses while employing different devices to interact with objects inside an immersive VE (Chittaro and Sioni, 2012a). The goal of this study is to highlight the importance of exploiting physiology in the evaluation of ergonomics of novel interaction devices. A large body of HCI research focuses on devices and techniques to interact with applications in more natural ways, such as gestures or direct pointing with fingers or hands. In particular, recent years have seen a growing interest in laser pointer–style (LPS) interaction, which allows users to point directly at the screen from a distance through a device handled like a common laser pointer. Several LPS techniques have been evaluated in the literature, usually focusing on users' performance and subjective ratings, but not on the effects of these techniques on the musculoskeletal system. One cannot rule out that “natural” interaction techniques, although found attractive by users, require movements that might increase likelihood of *musculoskeletal disorders* (MSDs) with respect to traditional keyboard and mouse. The study reported in this chapter investigates the physiological effects of a LPS interaction technique (based on the Wii Remote) compared to a mouse and keyboard setup, used in a sitting and a standing posture. The task (object arrangement) is representative of user actions repeatedly carried out with 3D applications. The obtained results show that the LPS interaction caused more muscle exertion than mouse and keyboard. Posture played also a significant role. The results emphasize the importance of extending current studies of novel interaction techniques with thorough EMG analyses.

3.1 Introduction

A large body of HCI research is devoted to create interaction techniques and devices that could allow users to interact with applications in more natural ways, such as gestures or direct pointing with fingers and hands. In particular, recent years have seen a growing interest in *laser pointer–style* (LPS) interaction, which allows users to point directly at the screen from a distance through a device which is handled like a common laser pointer.

Applications projected on a large screen might especially benefit from LPS interaction, becoming more intuitive and natural to use with respect to mouse and keyboard. The use of laser pointers (Figure 3.1) is common in public presentations (e.g., during a meeting or a lecture), but the functions provided by these devices are generally limited to projecting a red or green spot on the screen and, if the device communicates with the computer through Bluetooth or IR interfaces, scrolling through presentation slides.

In the literature, many approaches have been proposed to integrate these basic functions with more complex features, allowing users to control the on–screen cursor and GUI *widgets* (windows, icons, buttons, scrollbars, etc.) with a single laser pointer (Kirstein and Muller, 1998; Shizuki et al., 2006) or multiple laser pointers (Davis and Chen, 2002), using video cameras to track the laser spots on the screen and translate them into inputs to the application. Researchers have also replaced laser pointers with other devices for



Figure 3.1: A commercial laser pointer.

more flexibility, proposing various techniques based on LPS interaction (Bowman et al., 1997; Cheng and Pulo, 2003; König et al., 2007; Ouramdane et al., 2006b).

Evaluations of these interaction techniques and devices usually focus on users' performance and subjective ratings. The latter are used to evaluate perceived comfort and task difficulty (e.g., Elmqvist and Fekete, 2008; MacKenzie and Jusoh, 2001) and, more rarely, perceived fatigue (e.g., Douglas et al., 1999), showing that users generally like LPS interaction. However, these studies do not consider some essential variables: while the effects of traditional interaction techniques and devices such as keyboard and mouse on the musculoskeletal system have been investigated in depth (Sommerich et al., 2006), very little is known about natural interaction techniques that have been proposed in recent years. More specifically, one cannot rule out the possibility that these novel techniques, although found attractive by users, require movements that might increase users' likelihood of *musculoskeletal disorders* (MSDs) with respect to traditional keyboard and mouse. It should also be noted that MSDs develop insidiously over time: the user is typically asymptomatic and unaware of the negative effects of the movements she repetitively performs until it is too late.

For the above reasons, physiology–based ergonomics studies of natural interaction techniques are urgently needed and designers should be concerned about the future effects on users' health of the continuous use of the proposed techniques. In the present chapter, we thus investigate physiological effects of a LPS interaction technique compared to a traditional mouse and keyboard setup. In particular, we carry on a thorough analysis of on muscle activity, also considering the typical postures in which users operate the considered devices. The sitting posture is the most common when using mouse and keyboard; on the other hand, the standing posture seems to be convenient for LPS interaction with large displays because of the mobility afforded to the user. Therefore, the second goal is to study the effects of the two considered techniques in each of the two postures.

To focus on a task which is representative of actual user activities repeatedly carried out with commercial 3D software, we choose an *object arrangement task* (Chittaro et al., 2009), which is relevant in different popular 3D tools, ranging from 3D modeling software (e.g., *3ds Max* (Autodesk, 2013)) to game development tools (e.g., *Unity* (Unity Technologies, 2013)), for building VEs. It necessarily involves (i) 3D navigation (finding and reaching the place where a given object has to be positioned), (ii) object selection, and (iii) object manipulation, i.e., properly positioning, orienting and scaling the object.

This chapter is organized as follows. In Section 3.2, we review the literature on LPS interaction, then Section 3.3 describes the interaction techniques considered in our study, while Section 3.4 presents the details of the experiment. Section 3.5 and 3.6 respectively illustrate and discuss the obtained results.

3.2 Laser Pointer–Style Interaction in the Literature

Several approaches to support pointing, selection and object manipulation based on LPS interaction have been proposed in the literature. In Section 3.2.1, we describe the solutions based on the *ray casting* tech-

nique, which uses a virtual light ray to grab objects, with the ray direction specified by the user's hand or an handheld device (Bolt, 1980; Mine, 1995; Bowman and Hodges, 1997), while Section 3.2.2 illustrates the results of user studies which have dealt with LPS interaction.

3.2.1 Techniques based on ray casting

Systems based on the ray casting technique can be classified in two categories based on the kind of light they exploit, i.e., systems that use visible lasers and systems that use infrared instead of visible light. We briefly analyze both categories.

Systems based on visible laser In these systems, video cameras are used to track a laser spot for matching the position of the on–screen cursor with the pointed location (Kirstein and Muller, 1998), also adding visual feedback to both the laser spot (echoing it with a different cursor image for each action performed) and the GUI widgets, in order to make evident where the interaction will happen (Olsen Jr and Nielsen, 2001). The laser spot color, shape and position are also used to trigger actions such as moving forward and backwards in a presentation or drawing on a virtual canvas (Shizuki et al., 2006).

Systems based on IR light IR cannot be seen by the human eye. Thus, pointing an IR light towards the screen does not mask on–screen information. This also facilitates the use of on–screen cursors (the visual feedback), which can be changed in a very flexible manner (König et al., 2007). However, the feedback loop between the user's motor and perceptual systems is less effective with invisible light than with visible lasers, because users can only see the on–screen cursor, which usually suffers from latency problems (Cavens et al., 2002). Some researchers hide even the on–screen cursor, so the problems of *hand jitter* (the normal hand tremor which, albeit small, might result in a more noticeable cursor tremor) and high latency cannot be visually perceived. (Cheng and Pulo, 2003) proposed, for example, to exploit *hotspots* (i.e., logical highlighting of the currently pointed object) to let users know where they are pointing.

The Wii Remote is an affordable and widespread IR–based pointing device, which exploits two arrays of IR lights (the *Sensor Bar*, Figure 3.2b) placed over or below the screen to detect the pointed location (see Section 3.3.1). A specular approach is taken by Matveyev and Göbel (2003), who positioned the IR camera above the screen projector, and placed a single IR light on the handheld device, subdividing it into three smaller rays. By calculating angles and distances among the resulting multiple spots (detected by an IR video camera placed behind the screen), they allow users to rotate and translate objects along the *z* axis. Moreover, users can mechanically change the distance of one IR spot from the other two, and the change is recognized as a select operation.

3.2.2 User studies

Evaluations of LPS interaction techniques have focused on the assessment of user performance in pointing, selection and object manipulation tasks performed on 2D interfaces and VEs. Studies of LPS interaction which focus on 2D interfaces are quite common, and typically require users to point at and select geometric shapes or GUI widgets which lie on a plane. These studies typically use standard (ISO 9241-9) tests for the evaluation of pointing devices (e.g., Oh and Stuerzlinger, 2002) and similar evaluation instruments (e.g., Olsen Jr and Nielsen, 2001; Myers et al., 2002; Campbell et al., 2008). *Tapping tests* are a typical example (e.g., Myers et al., 2002): participants are asked to move the cursor from a starting position and select a target figure, whose form, size and distance from other similar geometrical shapes are previously defined. The use of standard tests allows one to compare the results obtained with similar analyses performed on trackball, joystick, mouse and touchpad (Douglas et al., 1999; MacKenzie et al., 2001).

User studies of LPS interaction in VEs are less frequent. Bowman and Hodges (1997) evaluated two ray casting techniques, three virtual arm techniques based on *Go–Go* (Poupyrev et al., 1996), which allows a 3D virtual arm projected on screen to stretch longer than the user's real arm for reaching distant objects, and an indirect virtual arm stretching technique in which the buttons on a 3D mouse are used to stretch and retract the arm. Users could freely point, select and manipulate objects inside a VE using each technique and their comments were recorded to assess the strengths and weaknesses of each approach. Ray casting

turned out to be the best technique for object selection, while virtual arm techniques were better suited for object manipulation.

Ouramdane et al. (2006a) evaluated ray casting and Go-Go with and without the assistance of their *FOLLOW-ME* technique. With *FOLLOW-ME*, when the cursor is far from the object that has to be selected, speed of approach is equal to hand speed, while it is reduced when the cursor is near the object. In the close proximity of the object, there is only one degree of freedom, so users' movements are interpreted only as *move closer to* or *move away from* the center of the object and the cursor moves on a 1D curve. In the evaluation, users had to select a series of static or moving 3D target objects, randomly appearing inside a VE. Elapsed time between the selection of two target objects, as well as the distance between the virtual tool (virtual pointer or virtual hand) and the target object were both considered as performance indicators. Results showed that *FOLLOW-ME* assistance could decrease the selection time necessary to reach the target object when using Go-Go, while it made little difference when using ray casting, perhaps because this is already a good method for selection, as shown by Bowman and Hodges (1997).

Mulder (1998) found that ray casting could also be a good approach for an object translation task when used inside a VE projected on a *Cave Automatic Virtual Environment* (CAVE) system (Cruz Neira et al., 1992). Users were asked to move a 3D sphere from a starting position to a target position inside a virtual box. Seven translation techniques were tested using a 6 DoF wand: three position control techniques, in which the wand movements control the position of the sphere; two velocity control techniques, in which the wand position and orientation controls the movement speed of the sphere; and two mixed techniques. *Stick*, a position control technique in which the 3D object is virtually attached to the wand with a segment, produced the fastest translations. Moreover, users perceived *Stick* as easy to use, intuitive and not fatiguing.

Existing studies of natural interaction techniques do not generally include the evaluation of users' physiological responses. A notable exception is the stress assessment carried out by Bérard et al. (2009). The study measured the performance of four devices used as indirect pointing devices in a 3D placement task: a traditional mouse; a mouse used together with the *DepthSlider*, i.e., an optically tracked physical slider to control translations along the z axis; the *SpaceNavigator*, a commercial 6 DoF device used to perform translations inside the VE; an optically tracked Wii Remote). Three physiological signals (EDA, BVPA, HR) were recorded to measure the level of stress induced by using each device. Results showed that the mouse, despite lacking a third degree of freedom, was both more precise and less stressful than the other evaluated devices. Unfortunately, LPS interaction techniques were not included in this study, and the considered physiological recordings did not include users' muscle activity.

3.3 The Considered Interaction Techniques

Millions of Wii consoles have been sold, making this entertainment device very popular. The Wii primary interface device is the Wii Remote, which is an intuitive, widespread and very well known device not only among gamers, but also in the general public. Moreover, the Wii Remote is being used as an alternative to mouse and keyboard also in some PC applications¹. Our study employs the Wii Remote for the LPS condition, while the second condition is based on a traditional mouse and keyboard setup.

In this section, we first illustrate the Wii Remote and how it was used in our study, then we describe in detail how users navigate inside the VE, point at and manipulate objects in each of the two experimental conditions.

3.3.1 The Wii Remote

The Nintendo Wii Remote (Figure 3.2a) includes a small IR video camera with a resolution of 1024×768 pixels and a FoV of about 40° , capable of tracking up to four IR sources with a 100 Hz sampling. This allows to use a Sensor Bar (Figure 3.2b), which contains two arrays of IR LEDs and is typically placed above or under the screen, to support detection of the location pointed by the device as well as the roll angle and the distance between the device and the bar itself. An additional device called Nunchuck, which includes a joystick and two buttons, can be connected to the Wii Remote as shown in Figure 3.2a.

¹see, for example, the applications listed at <http://www.brianpeek.com/page/net-based-wiimote-applications.aspx>

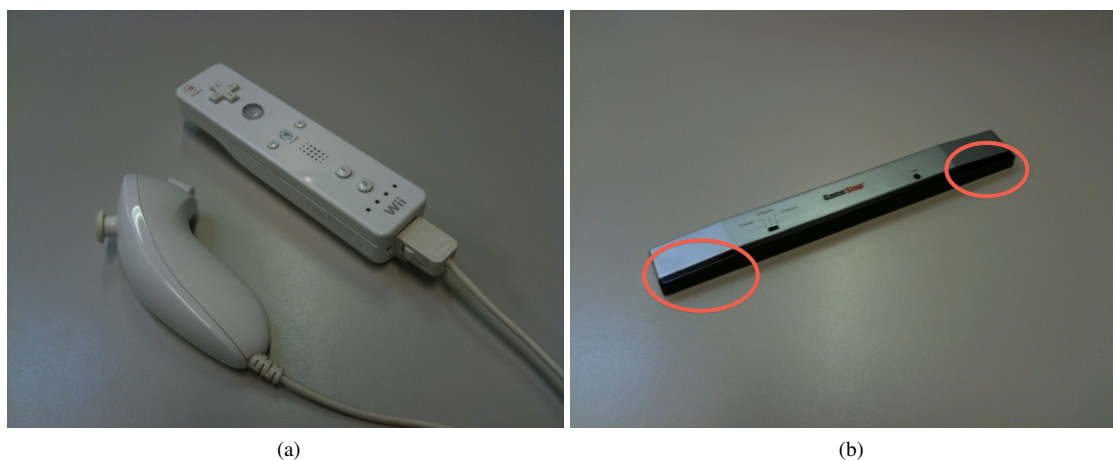


Figure 3.2: (a) The Nunchuck (on the left) connected to the Wii Remote (on the right); (b) the Sensor Bar used in the study (the two red circles highlight the position of the two IR LED arrays).

To read the output data sent by the Wii Remote through a Bluetooth connection, we employed the *WiiYourself!* (Gl.tter, 2007) library, that can return the position of each IR source relative to the IR camera viewport, the state of the buttons (neutral or pressed) and the tilt of the joystick on the Nunchuck.

To have the Wii Remote act as a direct pointing device that controls an on–screen cursor, the cursor coordinates are indirectly derived from (i) the position of the sensor bar relatively to the screen, (ii) the distance between the two IR LED arrays detected by the Wii Remote, and (iii) the size of the screen itself. If the difference between the actual pointed location and the cursor position on the screen is large, users might find it difficult to use the device, and the pointing performance might decrease (Cavens et al., 2002). To prevent this issue, we identified the Sensor Bar position and orientation for the experiment through a pilot test that preceded the user evaluation. Sensor Bar position and orientation were then checked before every user test to prevent possible deviations from the identified setup.

The smallest user’s movement that can be detected by the Wii Remote, i.e., moving the cursor position one pixel of the IR camera resolution, would translate in our case to a cursor movement of about 2 pixels on the projection screen (corresponding to about 2–3 mm on the screen; see Section 3.4.2). Therefore, even if the user tries to keep the Wii Remote still, her hand tremor would make the cursor jitter noticeable on a projection screen. To solve this issue, a smoothing algorithm was implemented: after dismissing small movements (resulting in cursor motions of less than 1 cm on the projection screen), it calculates the smoothed cursor position as the weighted mean of the last 10 filtered locations (the more recent the detected position, the greater the weight). Since this inevitably introduces lag, during the pilot test we fine–tuned the filter to minimize lag, while keeping jitter under a level considered acceptable by users.

To calculate the total latency for the two devices (i.e., not only the latency caused by filtering, as with the Wii Remote, but also caused by other factors, such as the wired and wireless connection with the PC and the buffering of the screen projector), we employed a technique inspired by the one described by Pavlovych and Stuerzlinger (2009). Using a video camera, we recorded at 50 frames per second (i) the projection screen, (ii) the mouse moving the cursor on the screen from side to side, and (iii) the Wii Remote moving the cursor on the screen from side to side. Both devices were lying on a desk and, for each one, the cursor was repeatedly moved on the screen for a couple of minutes, with an interval of about 2 s between each pair of subsequent motions. To make it easier to calculate the latency of the cursor with the Wii Remote, we attached a real laser pointer to the device which projected a laser spot over the actual pointed location. Total latency is the difference between the instant when the device stops and the instant when the cursor stops. The video was analyzed manually, and we averaged a total of 15 measurements to remove any potential sampling artifacts. The average latency between mouse and cursor movement was 76 ms ($SD = 8.3$), while it was unsurprisingly higher for the Wii Remote and equal to 233.3 ms ($SD = 20.9$).

3.3.2 Navigating the virtual environment

In a VE, objects can be far from users, thus perceptively small and difficult to select and manipulate with sufficient precision. Moreover, objects and locations might be outside the actual user's field of view. For this reasons, in an object arrangement task users should be able to move the viewpoint.

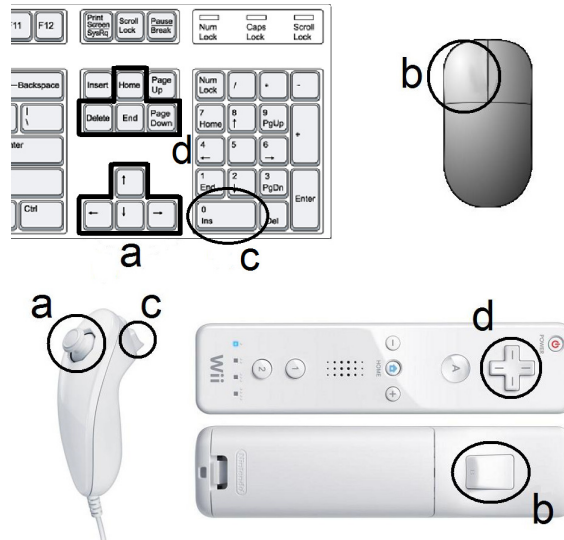


Figure 3.3: The controls for the two interaction techniques. *a*: navigation; *b*: viewpoint orientation, object selection and coarse manipulation; *c*: manipulation type; *d*: fine manipulation.

To navigate the VE in our study, participants used the arrow buttons on the keyboard or the joystick on the Nunchuck (see Figure 3.3a). Behavior of controls has been kept consistent with typical conventions adopted in video games. By pressing the up and down buttons on the keyboard (respectively, moving the joystick forward and backwards on the Nunchuck), users move the viewpoint forward and backward in the VE along their current viewing direction. By pressing left or right buttons on the keyboard (respectively, moving the joystick left or right on the Nunchuck), users change viewpoint orientation, rotating it counter-clockwise or clockwise respectively. In video games (especially *first person shooters*), left and right arrows or joystick movements are often used to *strafe*, i.e., move laterally the viewpoint without changing its orientation. However, this approach could be too complex for users who have little or no experience with video games. Our technique thus follows the approach of simpler video games, which allows for the viewpoint to move and change orientation by using only the four keyboard arrows or the Nunchuck joystick. Both motion speed and rotation speed are constant.

As an additional possibility, viewpoint orientation can also be changed using the mouse or the Wii Remote: the viewing direction can be moved up, down, left or right by pressing the left button on the mouse (respectively, the trigger *B* button on the Wii Remote, see Figure 3.3b) when the cursor is not over an object, and then dragging the cursor over the screen.

3.3.3 Pointing at and selecting an object

To point and select with LPS interaction, we employed IR-based ray casting. In particular, we have been inspired by other systems proposed in the literature (e.g., Olsen Jr and Nielsen, 2001; König et al., 2007) and also from some video games for the Wii console, e.g., Konami's *Eledees* (Konami, 2007) in which players have to make small creatures come out from objects inside a house by grasping, tossing and shaking those objects.

The on-screen cursor moves along the x and y axis, and when it hovers over an object, a *bounding box* for the object is highlighted (see the white box around the large “K” object in Figure 3.6). This approach is used, for example, by Cheng and Pulo (2003). Then, by pressing the left button on the mouse (respectively,

the “B” button on the Wii Remote, see Figure 3.3), the object is *selected* (i.e., it is possible for the user to manipulate it). We employed the “B” button on the Wii Remote and the left mouse button, already used to change viewpoint orientation, because these buttons are easily accessible to users. The pilot test showed that using the same button for object selection and for viewpoint orientation did not generate confusion in users.

Since the experimental task concerned the arrangement of objects which stand on the floor, to facilitate users in carrying out manipulations we constrained objects to be bound to the VE floor (in other words, they could not float above the floor or on the walls). When an object is selected, the on-screen cursor is moved to the center of the object bottom surface (i.e., the contact surface between the floor and the object). The cursor remains attached to the center of the bottom surface until the user releases the button, then it returns to the location currently pointed by the mouse or the Wii Remote. During the pilot test, we also tried the approach of always keeping the cursor in the position pointed by the device, but users preferred the solution described above, because the cursor better highlights the actual position of the object on the floor, so they felt more confident in performing object translations.

3.3.4 Manipulating the selected object

Once an object is selected, three *manipulation types* can be applied to it, i.e., the user can change its position (by moving it around), orientation (by rotating it around its vertical axis) and size (by making it bigger or smaller). Only one manipulation type can be performed at a time to keep controls as simple as possible. Using the “0” key on the numeric keyboard (respectively, the “Z” button on the Nunchuck, see Figure 3.3), the user can switch among the three manipulation types.

Each manipulation type is associated to a different cursor icon on the screen, giving visual feedback about which manipulation can be currently applied (Figure 3.4). The cursor icon is normally white, but it turns green while manipulation is performed.



Figure 3.4: From left to right, the three cursor icons associated respectively to translation, rotation and size manipulation.

After selecting an object, the user drags the cursor on the screen to manipulate it. Dragging changes one of the object properties (position, orientation or size) based on the chosen manipulation type.

With translation, the user changes the object position inside the VE. The vertical component of the cursor drag moves the object closer or away from the user, while the horizontal component moves the object to the left or to the right.

During an object arrangement task, users often need to move the selected object to a position that is outside the current field of view. Generally, these situations can be handled by exploiting the navigation controls (respectively, keyboard arrows and Nunchuck joystick) during an object translation. However, to make it easier to change the orientation of the field of view, we allow users to simply move the object near the screen borders and the viewpoint orientation inside the VE continuously changes, while maintaining the object inside the user’s field of view.

With rotation, the user changes the orientation of the object along its vertical axis. The user rotates the selected object clockwise by dragging the cursor to the left, and counter-clockwise by dragging the cursor to the right. The angle of object rotation is proportional to the length of the cursor drag.

With size manipulation, the user changes the object size. By dragging the cursor up and down, the object becomes respectively bigger or smaller. The change in size is proportional to the length of the cursor drag.

In the following, we refer to all the above described manipulation capabilities as *coarse manipulation*. Besides coarse manipulation, we provide also *fine manipulation* capabilities respectively through the four buttons on the keyboard and the *directional pad* on the Wii Remote (the “plus”-shaped button) as highlighted in Figure 3.3. Fine manipulations occur at a low, fixed speed to give users the ability to manipulate

objects more accurately. More precisely, a fine translation moves an object for about 0.6 m in a second inside the VE (which measures 7.5 m × 13.3 m), a fine rotation rotates an object for about 70° in a second, and a fine size manipulation makes an object about 60% bigger or smaller in a second. In coarse manipulation, each user can instead perform manipulations at a different variable speed, which can reach values as high as about 10 m in a second inside the VE for translations and 500° in a second for rotations. Due to this variability in performing coarse manipulations, the ratio between the speed of coarse and fine manipulations varies with the user.

Fine and coarse manipulations are mutually exclusive and, to prevent errors, the controls for coarse manipulation are disabled when using the controls for fine manipulation, and vice versa. Participants were thus invited to use one hand for coarse and fine manipulations, and the other hand to perform navigation and choosing the manipulation type.

In fine manipulations, the controls that move the selected object close or away from the users are respectively the two central keyboard keys among the four highlighted in Figure 3.3 and the *up* and *down* buttons of the directional pad on the Wii Remote, while the controls that move the selected object left or right are the other two keyboard keys in Figure 3.3 and the *left* and *right* buttons of the directional pad on the Wii Remote. During a fine rotation, the selected object can be rotated clockwise or counter-clockwise by pressing respectively the leftmost and rightmost keyboard keys among the four highlighted in Figure 3.3 and the *left* and *right* buttons of the directional pad on the Wii Remote. During a fine size manipulation, the object can be made bigger or smaller by pressing respectively the two central keyboard keys in Figure 3.3 and the *up* and *down* buttons of the directional pad on the Wii Remote.

3.4 Experimental Evaluation

The study follows a within-subject design with *interaction technique* (based on *Wii Remote and Nunchuck* or based on *mouse and keyboard*) and *posture* (*sitting* or *standing*) as the independent variables (IVs). For conciseness, in the following we refer to Wii Remote and Nunchuck as *W&N*, and to mouse and keyboard as *M&K*.

3.4.1 Participants

The evaluation involved a sample of 18 users (13 M, 5 F) with various educational backgrounds (7 computer science, 2 literature and philosophy, 2 engineering, 2 physiotherapy, 1 education, 1 mathematics, 1 agricultural science, 1 mechanical design, 1 natural science), recruited among graduate and undergraduate university students and people from other occupations. Their age ranged from 20 to 58 ($M = 27.4$, $SD = 8.8$).

All of them had at least basic experience with M&K and 11 of them used M&K with VEs (mostly games) at least once a week. 5 participants used W&N at least once a week to play games.

3.4.2 Materials

The VE was run on a PC in fullscreen mode and projected on a 240 × 160 cm projection screen at UXGA resolution (1600 × 1200 pixels). The distance between the screen and the user was about 3.5 m. A common USB mouse and PS/2 keyboard were employed for M&K.

To record users' physiological data, we employed seven sensors, positioned as shown in Figure 3.5: four EMG sensors, coupled with disposable triode electrode pads; a thermometer for peripheral temperature (recorded on the little finger of the left hand); a PPG for BVP (recorded on the ear lobe); a girth sensor for respiration measurements. These signals were recorded and stored on a second PC. Two webcams were used to record videos of respectively the projection screen and the body of the subject during the task for reviewing purposes.

During the experiment, two VEs were used. A training VE represented a group of three houses and was used by participants to practice with controls. The experimental VE reproduced a rectangular room of a museum and was used to perform the experimental task. The frame rate of the projected image was kept above 30 frames per second for both VEs.

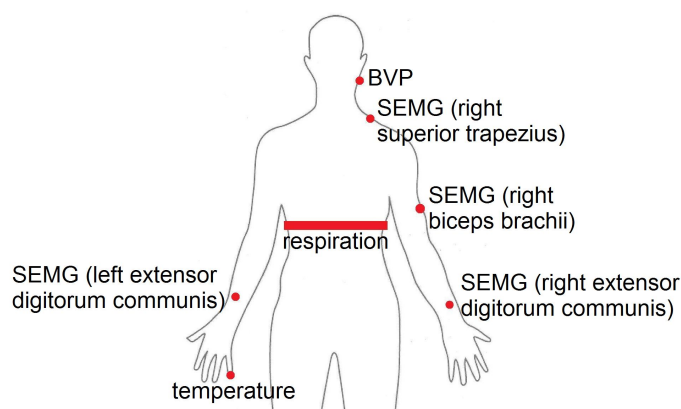


Figure 3.5: Position of the seven sensors. The silhouette represents the back of a human body.

3.4.3 Task

For each of the four combinations of interaction technique and posture, participants performed a task which requires to arrange a blue “K” object and a green “Z” object. Each of the two objects was initially positioned in front of the user and not selected. Users had to arrange each object in such a way that it matched position, orientation and scale of a red, semitransparent copy of the object itself (see, for example, Figure 3.6).

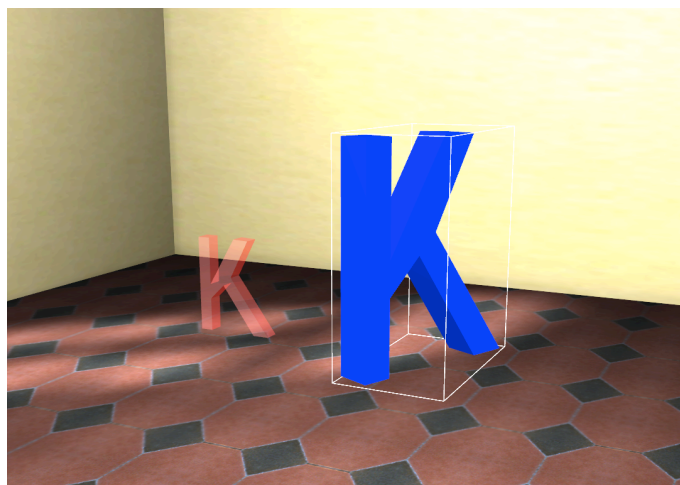


Figure 3.6: An object to be arranged (on the right) and the corresponding target object (on the left). In this example, the object to be arranged is currently selected (bounding box highlighted).

For each experimental condition, object position and orientation as well as initial user orientation were changed. However, to keep task complexity constant, objects and targets were always placed at the four corners of the museum room and initial distances as well as differences in size and orientation between object and target remained the same. Users started always at the center of the VE, facing the two objects they had to arrange during the task (Figure 3.7a).

The task started with a short sound, which was repeated as soon as the object and the corresponding target were properly matched (Figure 3.7b). Error tolerance thresholds were used in determining object and target match: a match was detected when (i) the distance between object and target is no more than 0.2

times the height of the target; (ii) the size of the object is between 80% and 120% the size of the target; (iii) the difference in orientation between the object and its target is no more than 30° . These thresholds were determined during the pilot test to obtain a level of complexity that is reasonable for users. The system did not check the accuracy of the match while users were carrying out a manipulation, but only after its conclusion, to prevent participants from getting the correct match by chance by simply performing quick object manipulations over targets.

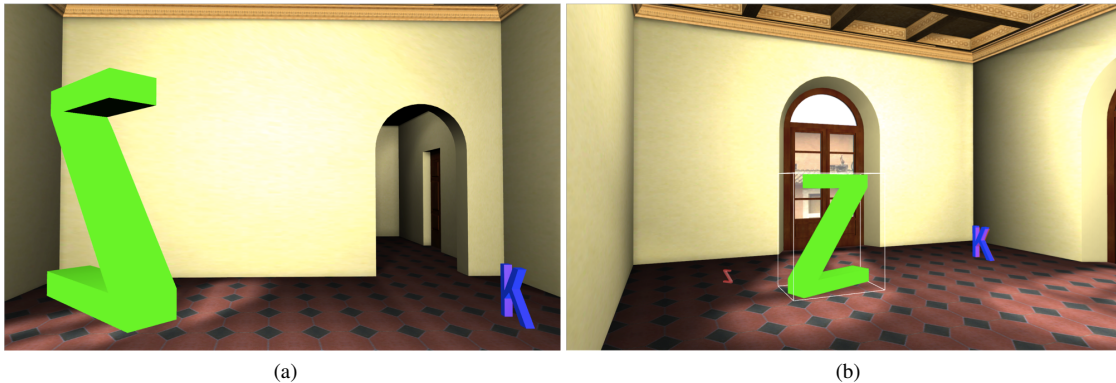


Figure 3.7: (a) User's viewpoint at the beginning of the task; (b) the user has matched the "K" target, and is arranging the "Z" object.

3.4.4 Procedure

Participants were verbally briefed about the nature of the task, the use of the physiological sensors and the use of the two webcams during the test. They were asked to fill a demographic questionnaire concerning gender, age, educational background and occupation, experience with MK and WN, experience with 3D video games and 3D software.

All participants chose to use keyboard and Nunchuck with the left hand and Wii Remote and mouse with the right hand, including the only left-handed subject. We thus measured the biceps and trapezius activity from the right arm and shoulder as illustrated in Figure 3.5.

The skin of forearms, right arm and right shoulder of participants was cleaned using a cotton pad and denatured alcohol, then the seven sensors were applied (see Section 3.4.5). A sheet with a human silhouette (Figure 3.5), pinpointing the exact position of each sensor, was shown to participants to let them know in advance how many sensors had to be applied and where. To facilitate the placement of the four SEMG electrodes, participants were asked to wear only a T-shirt during the experiment. Room temperature was maintained at about 21°C for reliable measurement of participants' peripheral temperature.

Once the sensors were placed, participants sat in a comfortable position and were asked to relax for about three minutes, while the baseline for the physiological signals was recorded. During this time, a video with relaxing images and music was shown in a dim light. Participants could close their eyes and only listen to the music if they preferred.

Each participant then performed the task in the four conditions following a different order, to prevent learning effects. The order of the four different object placements (see Section 3.4.3) was varied independently. There were 24 possible orders of conditions and 24 orders of object placement. Each participant was randomly assigned to one order of condition and one order of object placement in such a way that each order was assigned to at most one participant.

To comfortably use MK, a common desk (78 cm tall) and a taller one (106 cm) were employed, respectively for the sitting and standing posture. In the sitting condition, the chair was adjustable in height and equipped with armrests, which could be used by participants as they felt more natural (Figure 3.8).

Before each task, participants were allowed to spend unlimited time trying an object arrangement task inside the training VE to practice with controls and familiarize with the task itself and the tolerance levels



Figure 3.8: A participant performing a task with W&N in sitting position.

of object and target match. During this time, they were provided with an instruction sheet illustrating the controls of each interaction technique; they could also ask questions to the experimenter. After training, participants were asked to wait for about three minutes to return to a relaxed state.

During the experimental task, the control instruction sheet, as well as a simple map of the experimental VE, pinpointing the participants' starting position as well as the position of objects and targets (Figure 3.9), were available to participants. The map was used to make participants initially aware of the position of the two targets, which were not visible from the starting position.

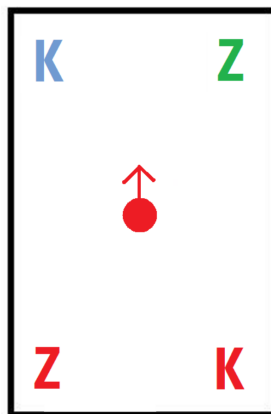


Figure 3.9: An example of room map given to participants. The participant's starting position as well as viewing direction are represented by the central dot and arrow. Colored letters show the position of the objects to be manipulated (green and blue letters, at the top of the map) and the targets (red letters, at the bottom). Note that the targets are initially outside the users' field of view.

After all tasks were completed, physiological sensors were removed and participants were asked to fill a questionnaire that asked to rank from 1 (best) to 4 (worst) the four conditions (ties were allowed) with respect to ease and comfort of navigation, ease and comfort of coarse and fine manipulation, perceived level of exertion and overall pleasantness of the condition.

3.4.5 Data collection

During the experiment, the following data were recorded:

- *Task completion time*: the time taken to complete the task, defined as the time elapsed between the starting sound and the detection of a correct match for both objects;
- *Navigation time, coarse manipulation time and fine manipulation time*: the time spent by users on each of the three activities;
- *BVP signal, respiration signal and temperature signal*: the signals recorded by the PPG, girth, and thermometer sensors;
- *surface EMG signals (left extensor digitorum communis, right extensor digitorum communis, right superior trapezius and right biceps brachii)*: the electric activity of these muscles, recorded on the participants' skin;
- *Subjective preferences*: the results of the ranked choice questionnaire.

From the raw physiological data just described, these additional data were derived:

- *HR, BVPA, RR, RA and peripheral temperature*: these values were averaged over 5-s epochs to reduce artifacts, especially BVP artifacts caused by head movements;
- *Mean activity of left extensor digitorum communis, right extensor digitorum communis, right superior trapezius and right biceps brachii*: the mean value of the RMS transformation of the four surface EMG signals, averaged over 1-s epochs (epochs for surface EMG signals need to be shorter than BVP epochs because of the faster variations of these signals). A *notch filter* (band-stop filter), centered on the 50 Hz frequency, was applied to remove typical AC interference caused by electronic devices;
- *Total muscle activity*: the value of the *integrated EMG* (IEMG) signals for each muscle. These values are derived from the area below the RMS EMG curves;
- *Mean values of the EMG power spectrum mean frequency*: each muscle fiber “discharges” at a particular frequency, so the EMG power spectrum (in which squared amplitudes of the frequency spectrum are considered) is the combination of the discharge frequencies of all the muscle fibers under the electrode. In the literature, the median frequency of the EMG power spectrum is sometimes measured, but it shows greater variability (Merletti et al., 1990; Andreassi, 2007). Values are smoothed over 2-s epochs;
- *EMG gradients*: linear regression slopes of each considered muscle activity, derived from the RMS transformation of the four surface EMG signals averaged over 1-s epochs;
- *Linear regression slopes of the mean values of the EMG power spectrum mean frequency*: the mean frequency trend over time for the considered muscle activities, derived from the mean frequency values of the power spectrum averaged over 2-s epochs.

The four studied muscles were chosen after consulting with an occupational therapy clinician and two physiotherapists. In particular, we measured the activity of (i) left and right extensor digitorum communis muscles because they extend the medial four digits of the hands (the two surface EMG sensors applied over the forearms are affected also by the activity of the thumbs), (ii) superior trapezius muscle of the arm used to hold the mouse and the Wii Remote, because this muscle is typically under heavy load in computer work (Wahlstrom, 2005) and (iii) biceps brachii muscle of the arm used to hold the mouse and the Wii Remote, because this muscle is particularly involved in the lifting of objects (the Wii Remote in our case). We decided not to focus on trapezius and biceps muscles of the other arm because they are much less activated by the considered task. As we reported in Section 3.4.4, all participants used the right arm to handle the mouse and the Wii Remote. Mean EMG and IEMG values were used to assess how much muscle effort was required to participants in each condition. EMG gradients were instead considered because there

is evidence that they are related to level of motivation: the higher the motivation, the steeper the slope (Andreassi, 2007).

Mean frequency values and their linear regressions are good indicators of a sustained muscle contraction and a signal of localized muscle fatigue (Andreassi, 2007; Merletti et al., 1990; Tassinari et al., 2007): the “faster” fibers contract at the beginning of muscular contraction and are subsequently replaced by the “slower” ones. Therefore, during a prolonged muscle activity the mean frequency decreases (the power spectrum graph tends to “shift” to the left).

Circulatory and respiration system measurements were used to assess the users’ stress level. Sympathetic arousal tends to increase heart rate and respiration frequency, while decreasing peripheral temperature, BVP amplitude and respiration amplitude (Andreassi, 2007). However, we also have to take into account that postural differences cause significant variations in the circulatory system (Jones et al., 2003): in particular, the standing position causes the heart rate to increase and the BVPA to decrease.

Task completion time as well as all key, button and joystick actions were automatically logged by the system. Action timings were used to calculate the duration of navigation and each manipulation type for each task.

3.5 Results

One participant was excluded from the analysis, because a 3D modeling error in the VE caused him to get stuck in a wall, preventing him to complete the first task. The sensor on the right forearm of a second user produced abnormal measurements of right extensor digitorum communis due to thick hair; therefore, values recorded on this muscle for this user were excluded from analysis.

Baseline values recorded before task execution (as described in Section 3.4.4) were removed from physiological data before analysis to account for individual differences.

To analyze data, we performed a repeated measures two-factor *analysis of variance* (ANOVA). Since this parametric test assumes that data follow a Gaussian distribution, we checked data normality using the Shapiro and Wilk (1965) normality test. When data were not normally distributed, we first tried to apply mathematical transformations to make the distribution more symmetric (Cohen, 2001). When data could not be normalized, we employed the non-parametric *ANOVA-type statistics* (ATS) proposed by Akritas and Brunner (1997) and further refined by Brunner et al. (1999).

When the analysis revealed interactions among IVs, we investigated them, as suggested by Cohen (2001), by performing cell-to-cell comparisons and adjusting the p value with Bonferroni correction. The comparisons were carried out using t -tests or, when the data could not be normalized, the non-parametric Wilcoxon test.

Overall, muscle activity results and completion times produced the most interesting outcomes, with M&K producing better results than W&N in most of the cases. This was reflected also in subjective preferences.

In the following sections, we describe all the findings in detail.

3.5.1 Task completion time

Task completion time data was not normally distributed and a square root transformation was applied to normalize it. Figure 3.10 shows the untransformed mean values of task completion time in the four experimental conditions. There was neither a significant interaction between the two IVs, nor a significant main effect of posture. A significant main effect of interaction technique was instead detected ($F(1, 64) = 12.27$, $p < 0.01$): carrying out the task with W&N required about 36% more time than with M&K.

3.5.2 Navigation time, coarse manipulation time, and fine manipulation time

Navigation time data in seconds was not normally distributed and a square root transformation was applied, while navigation time expressed as a percentage of the sum of navigation time, coarse manipulation time and fine manipulation time followed a Gaussian distribution. Figure 3.11 shows the untransformed mean values of navigation time expressed in seconds as well as a percentage. The analysis of navigation time in

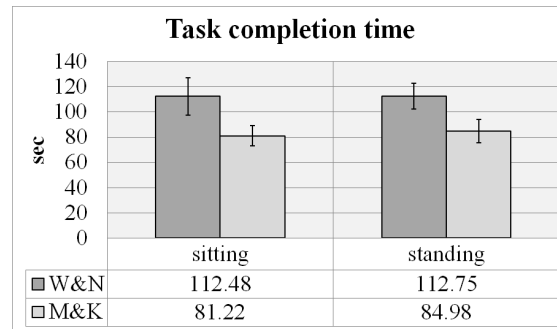


Figure 3.10: Mean completion time. Error bars indicate standard error of the mean.

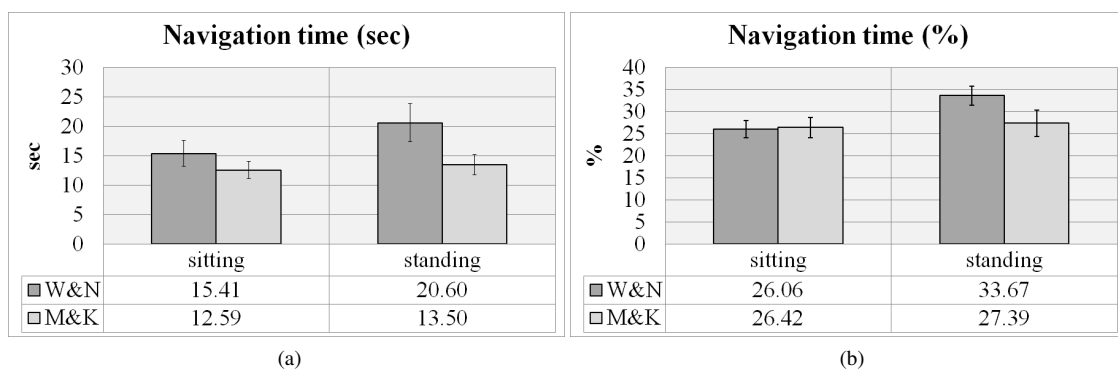


Figure 3.11: (a) Mean navigation time in seconds and (b) mean navigation time expressed as a percentage of the sum of navigation, coarse and fine manipulation times. Error bars indicate standard error of the mean.

seconds showed that interaction between the two IVs as well as main effect of posture were not significant. Main effect of interaction technique was instead significant ($F(1, 64) = 4.67, p < 0.05$): users navigated for a longer time with W&N than M&K. When analyzing the navigation time data as a percentage, neither interaction between the two IVs nor main effect of interaction technique were significant. Main effect of posture was instead significant ($F(1, 64) = 5.33, p < 0.05$): users navigated for a longer time when standing than when sitting.

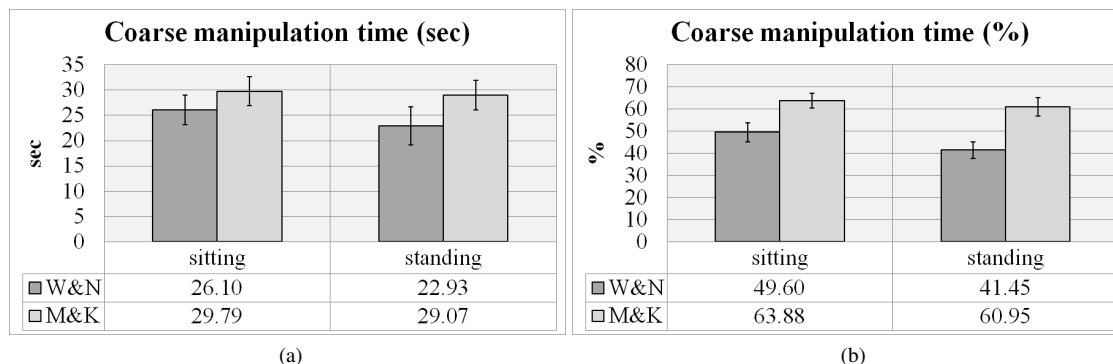


Figure 3.12: (a) Mean coarse manipulation time in seconds and (b) mean coarse manipulation time expressed as a percentage of the sum of navigation, coarse and fine manipulation times. Error bars indicate standard error of the mean.

Similarly to navigation time, coarse manipulation time data was not normally distributed and a square root transformation was applied. Figure 3.12 shows the untransformed mean values of coarse manipulation time expressed in seconds and as a percentage of the sum of navigation time, coarse manipulation time and fine manipulation time. Analysis of coarse manipulation time data in seconds showed lack of main effects and interaction. The analysis of coarse manipulation time data as a percentage showed neither a significant interaction between the two IVs, nor a significant main effect of posture. Main effect of interaction technique was instead significant ($F(1.64) = 16.77, p < 0.001$): users spent less time using coarse manipulation with W&N than M&K.

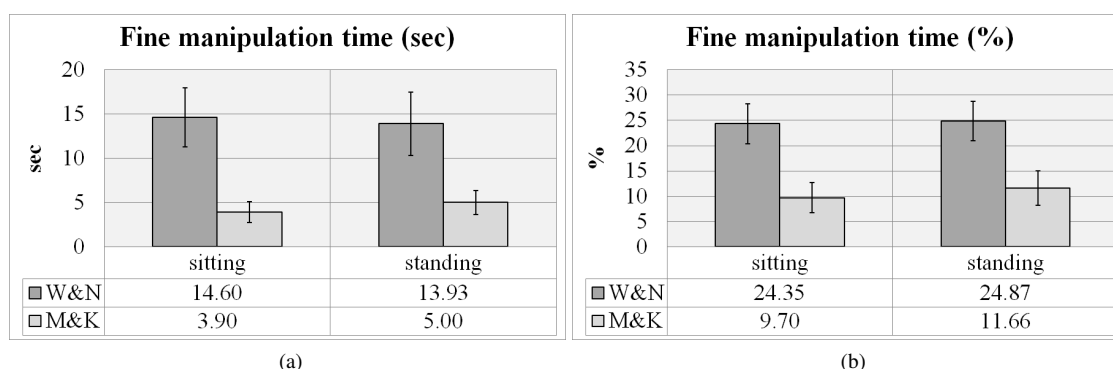


Figure 3.13: (a) Mean fine manipulation time in seconds and (b) mean fine manipulation time expressed as a percentage of the sum of navigation, coarse and fine manipulation times. Error bars indicate standard error of the mean.

Fine manipulation time data in seconds as well as a percentage could not be normalized. Figure 3.13 shows the mean values of fine manipulation time expressed in seconds as well as a percentage of the sum of navigation time, coarse manipulation time and fine manipulation time. The analysis of fine manipulation time in seconds revealed neither a significant interaction between the two IVs, nor a significant main effect of posture. Main effect of interaction technique was significant ($ATS = 27.40, p < 0.001$). The analysis of

fine manipulation time as a percentage produced similar results, with a significant main effect of interaction technique ($ATS = 19.09$, $p < 0.001$). These two results show that participants used fine manipulation for a longer time with W&N than M&K.

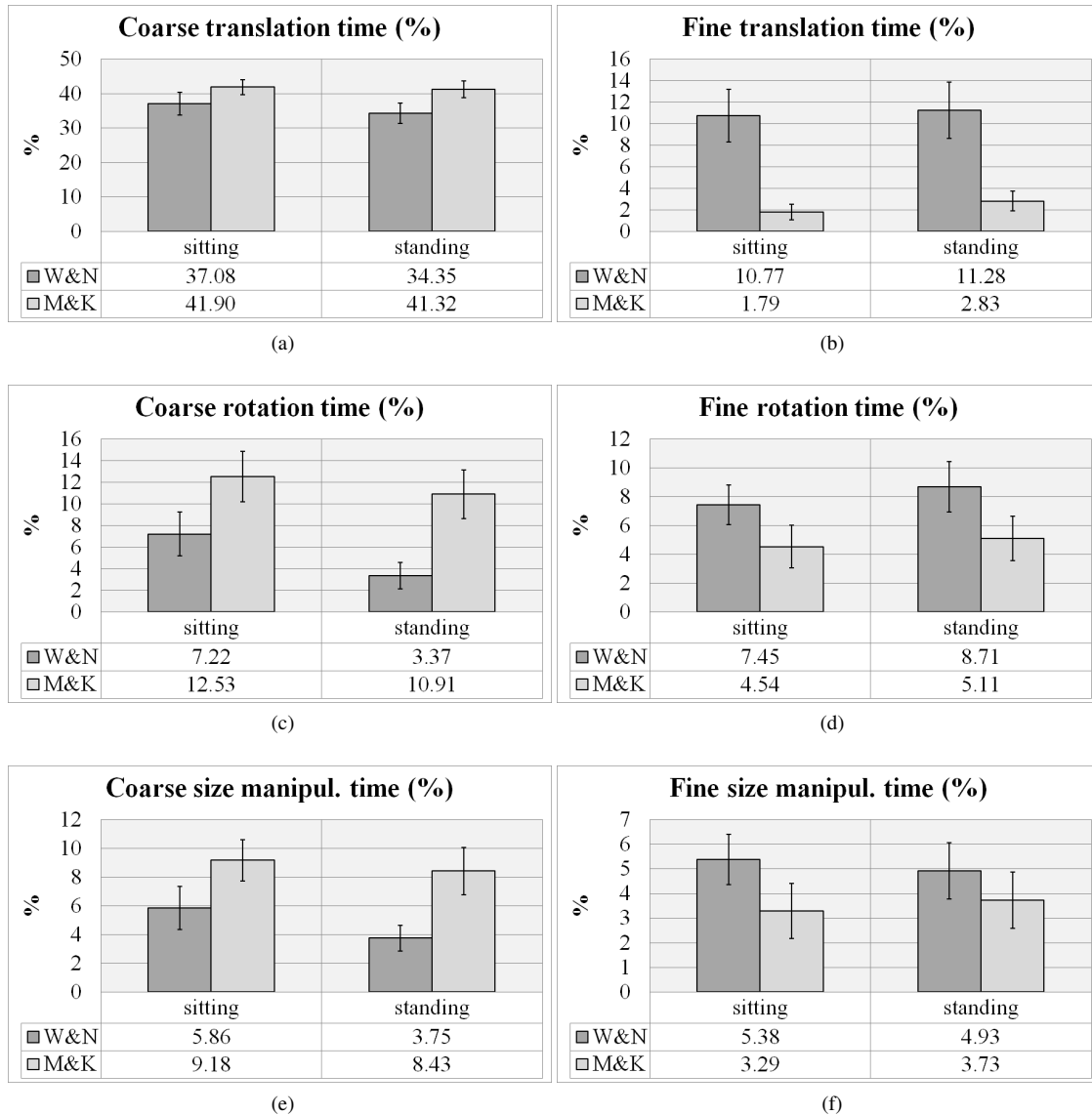


Figure 3.14: Mean time, in percentage of the sum of navigation time, coarse manipulation time and fine manipulation time, spent on coarse and fine translation, rotation and size manipulation. Error bars indicate standard error of the mean.

We analyzed in detail coarse and fine translation, rotation and size manipulation data expressed as a percentage of the sum of navigation time, coarse manipulation time and fine manipulation time. The mean values are shown in Figure 3.14. Coarse translation data was normally distributed, while other data were not following a Gaussian distribution and could not be normalized. The analysis showed no interaction between IVs, and five main effects: a main effect of interaction technique for fine translation time ($ATS = 32.58$, $p < 0.001$), coarse rotation time ($ATS = 10.38$, $p < 0.01$), fine rotation time ($ATS = 5.26$, $p < 0.05$), coarse size manipulation time ($ATS = 9.03$, $p < 0.01$) and fine size manipulation time ($ATS = 6.49$, $p < 0.05$). These results show that participants, with W&N, used less coarse rotation and size manipulation rather than M&K, while they used more fine translation, rotation and size manipulation with W&N rather than M&K.

Since, as seen in Section 3.3.2, participants could change viewpoint orientation in two different ways (i.e., by dragging the cursor with the mouse and the Wii Remote, and by pressing left and right arrow buttons on the keyboard and moving left and right the Nunchuck joystick), we analyzed in detail the time spent by participants using the two possibilities. The mean values are reported in Table 3.1. An ATS analysis of the difference in usage proportions of the two possibilities (see third column of Table 3.1) revealed no significant differences among conditions.

Task	Orient. with pointing (%)		Orient. with keys/mouse (%)		Difference (%)	
	Mean	SD	Mean	SD	Mean	SD
M&K sitting	26.09	29.34	73.91	29.34	47.83	58.68
M&K standing	34.33	32.09	65.67	32.09	31.33	64.19
W&N sitting	29.99	29.22	70.01	29.22	40.03	58.44
W&N standing	33.78	27.38	66.22	27.38	32.44	54.75

Table 3.1: Time spent by users in viewpoint orientation, split in percentage between usage of mouse and Wii Remote pointing, and usage of keyboard buttons and Nunchuck joystick. The third column provides the average difference of the two percentages.

3.5.3 Circulatory and respiratory activity and temperature measurements

If baseline values are greater than the recorded values during the task, the data obtained by subtracting the baseline are negative. In the following sections, when all mean values in a chart are negative, we reverse the scale for ease of reading.

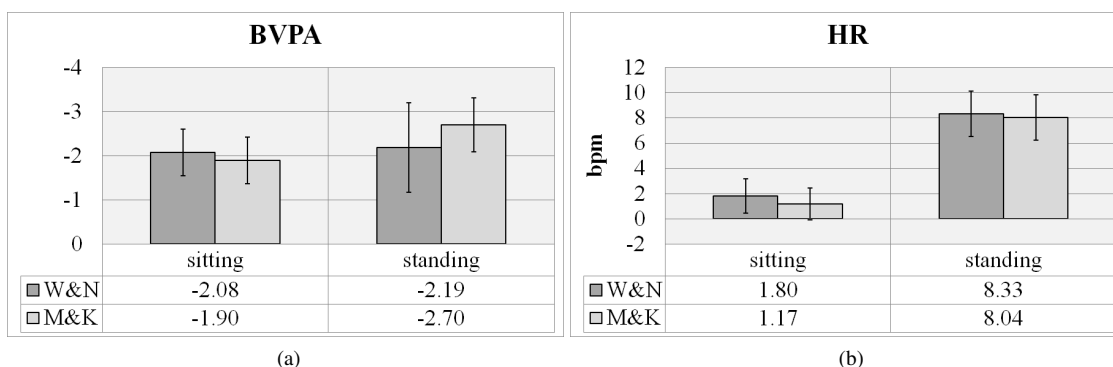


Figure 3.15: Change of BVPA (chart with reversed scale) and HR with respect to baseline values. Error bars indicate standard error of the mean.

BVPA data was not normally distributed and could not be normalized, while HR data followed a Gaussian distribution. Figure 3.15 shows the untransformed mean values for the two measures. The figure does not specify a measure unit for BVP amplitude, because amplitude concerns a relative quantity. ATS analysis of BVP amplitude revealed neither a significant interaction between the two IVs, nor a significant main effect of interaction technique. A significant main effect of posture was instead detected ($ATS = 7.83$, $p < 0.01$). Similarly, ATS analysis of heart rate revealed a significant main effect of posture ($F(1, 64) = 32.90$, $p < 0.001$): participants had lower BVPA values and higher HR with respect to baseline values when standing than sitting.

Respiration amplitude and respiration frequency data could not be normalized. Figure 3.16 shows the mean values for the two measures. The analysis of these two measures revealed neither a significant interaction between the two IVs, nor significant main effects of interaction technique and posture.

Peripheral temperature data was not normally distributed. All values were initially increased by a constant to become equal or greater than 1, in such a way that a square transformation could be applied. Figure 3.17 shows the untransformed mean values for this measure. The analysis revealed no significant interaction between the two IVs and no significant main effect of posture. Main effect of interaction technique was

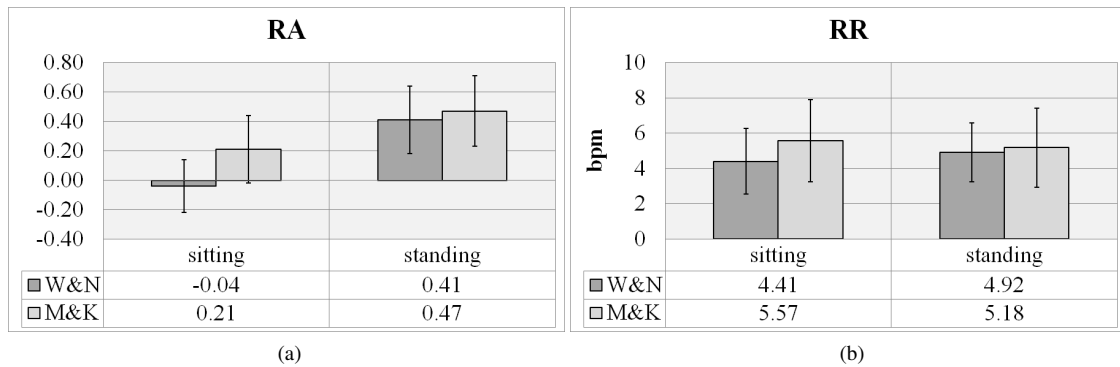


Figure 3.16: Change of respiration amplitude and respiration frequency with respect to baseline values. Error bars indicate standard error of the mean.

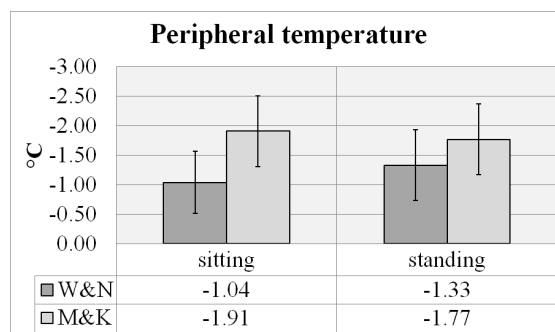


Figure 3.17: Change of peripheral temperature (chart with reversed scale) with respect to baseline values. Error bars indicate standard error of the mean.

instead significant ($F(1,64) = 8.77, p < 0.01$): participants showed a higher peripheral temperature with W&N than M&K with respect to baseline values.

3.5.4 Mean muscle activity

Left extensor digitorum communis and right trapezius data were initially not normally distributed. To obtain normality, these data were initially increased by a constant to become equal or greater than 1, so that a square root transformation could be applied to both data sets. Right extensor digitorum communis and right biceps brachii data followed a Gaussian distribution. Figure 3.18 shows the untransformed means for left extensor digitorum communis and the means for right extensor digitorum communis, while Figure 3.19 shows the untransformed means for right superior trapezius and the means for right biceps brachii.

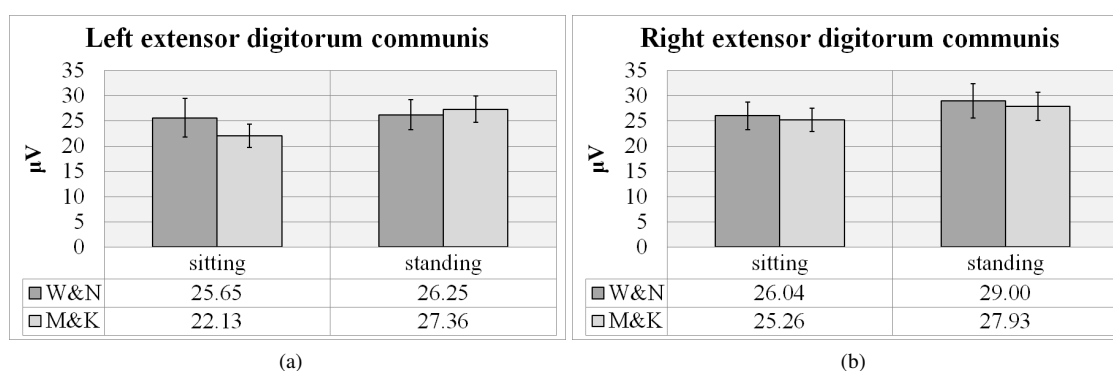


Figure 3.18: Change of mean activity for left and right extensor digitorum communis with respect to baseline values. Error bars indicate standard error of the mean.

The analysis of mean activity of left and right extensor digitorum communis revealed neither a significant interaction between the two IVs, nor a significant main effect of interaction technique. Main effect of posture was instead significant for both muscles (left muscle: $F(1,64) = 6.98, p < 0.05$; right muscle: $F(1,60) = 5.48, p < 0.05$): participants showed a greater mean activity of left and right extensor digitorum communis when standing than sitting.

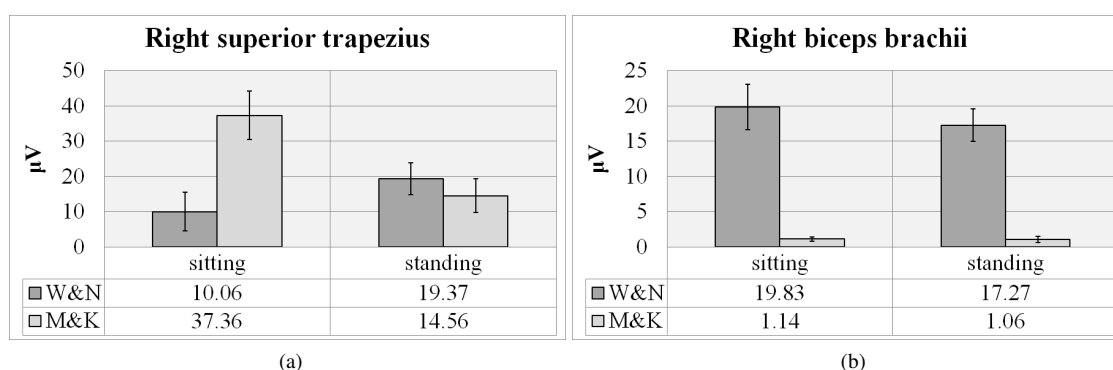


Figure 3.19: Change of mean activity for right superior trapezius and right biceps brachii with respect to baseline values. Error bars indicate standard error of the mean.

The analysis of right superior trapezius data revealed a significant interaction between the two IVs ($F(1,64) = 32.39, p < 0.001$). A significant main effect of interaction technique ($F(1,64) = 5.60, p < 0.05$) was also detected, while main effect of posture was not significant. Investigation of the interaction showed that with M&K the mean activity of right superior trapezius was significantly higher when sitting than standing ($t(16) = 3.92, p < 0.01$). Results show also that mean muscle activity was significantly higher

with M&K than W&N ($t(16) = 3.96, p < 0.01$) when sitting. No significant difference between W&N and M&K was found when standing.

The analysis of right biceps brachii data revealed neither a significant interaction between the two IVs, nor a significant main effect of posture. Main effect of interaction technique was significant ($F(1, 64) = 41.08, p < 0.001$): mean biceps muscle activity was higher with W&N than M&K.

3.5.5 Total muscle activity

IEMG data recorded from left extensor digitorum communis was normally distributed. To normalize right extensor digitorum communis and right superior trapezius data, a square root transformation was applied to the former, while the latter was initially increased by a constant to become equal or greater than 1, in such a way that a square root transformation could be applied. Right biceps brachii could not be normalized. Figure 3.20 shows the untransformed IEMG values for left and right extensor digitorum communis, while Figure 3.21 focuses on right superior trapezius and right biceps brachii.

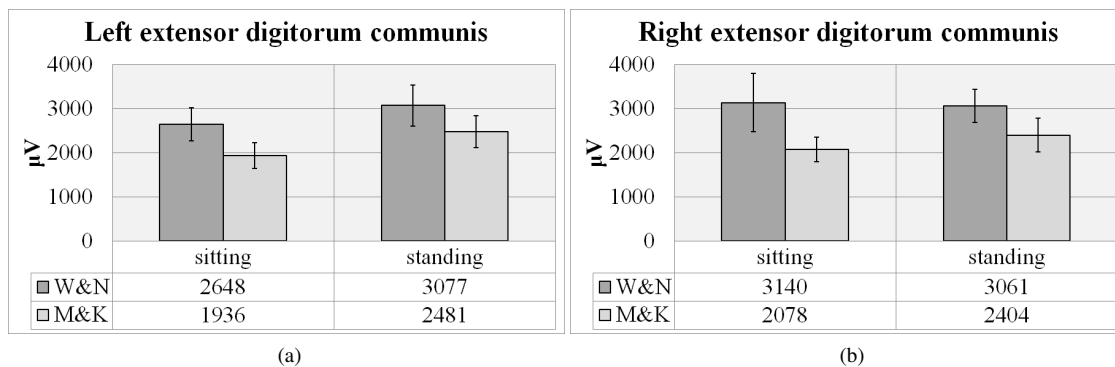


Figure 3.20: Change of IEMG data for left and right extensor digitorum communis with respect to baseline values. Error bars indicate standard error of the mean.

The analysis of IEMG data of left and right extensor digitorum communis revealed no interaction and one main effect. The main effect of interaction technique for the right extensor digitorum communis was significant ($F(1, 60) = 9.53, p < 0.01$): total muscle activity was higher with W&N than M&K.

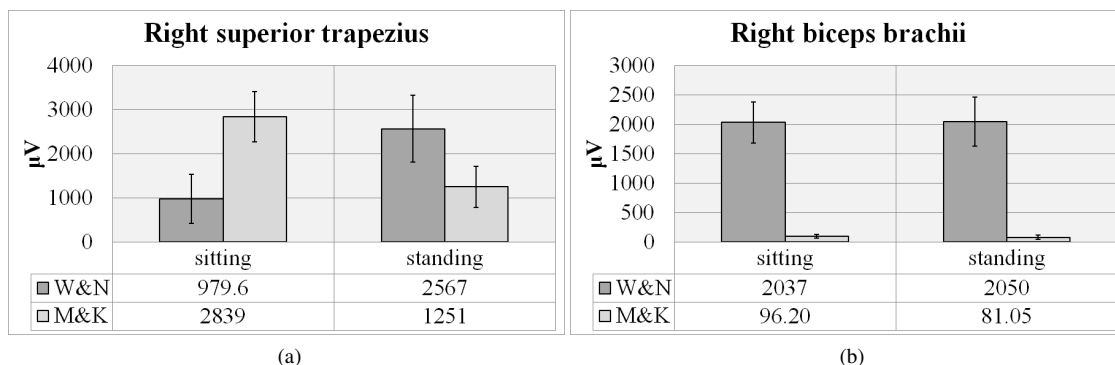


Figure 3.21: Change of IEMG data for right superior trapezius and right biceps brachii with respect to baseline values. Error bars indicate standard error of the mean.

The analysis of IEMG data of right superior trapezius showed that interaction between the two IVs was significant ($F(1, 64) = 26.39, p < 0.001$), while both main effects were not significant. Investigation of the interaction showed that with M&K the total activity of right superior trapezius was significantly higher when sitting than standing ($t(16) = 3.73, p < 0.01$). With W&N, total muscle activity was significantly

higher when standing than sitting ($t(16) = 3.27, p < 0.01$). Moreover, IEMG was significantly higher with M&K than W&N ($t(16) = 3.33, p < 0.01$) when sitting.

ATS analysis of IEMG data of right biceps brachii revealed no significant interaction between the two IVs and no significant main effect of posture. Main effect of interaction technique was significant ($ATS = 83.14, p < 0.001$): the total activity of the right biceps brachii was higher with W&N than M&K.

3.5.6 EMG power spectrum mean frequency

Mean frequency data from the activity of left extensor digitorum communis, right superior trapezius and right biceps brachii was normally distributed. It did not follow a Gaussian distribution and could not be normalized for right extensor digitorum communis.

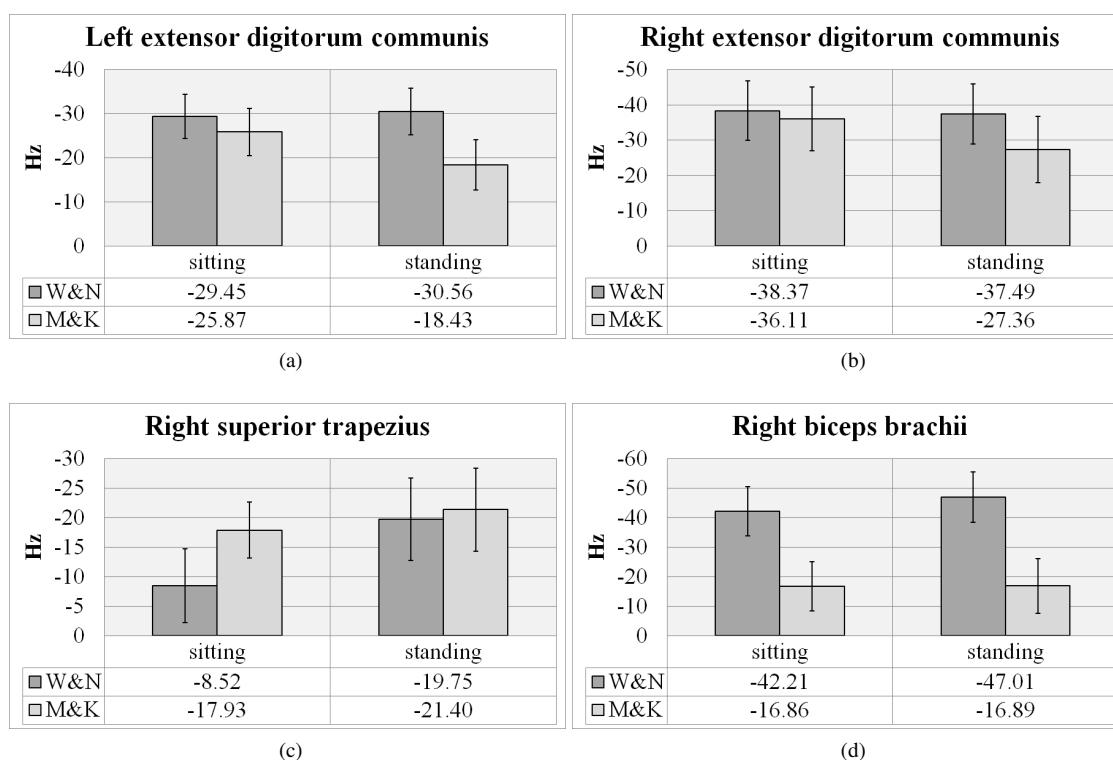


Figure 3.22: Change of mean power spectrum frequency for the four muscles (charts with reversed scale) with respect to baseline values. Error bars indicate standard error of the mean.

The analysis of mean frequency data for left extensor digitorum communis (Figure 3.22a) revealed a significant interaction between the two IVs ($F(1, 64) = 7.44, p < 0.05$) and a main effects of interaction technique ($F(1, 64) = 14.32, p < 0.01$) as well as posture ($F(1, 64) = 5.21, p < 0.05$). Investigation of the interaction showed that with M&K the mean frequency was higher when standing than sitting ($t(16) = 3.20, p < 0.01$), and that the mean frequency was higher with M&K than W&N ($t(16) = 5.01, p < 0.001$) when standing.

The analysis of mean frequency data for right extensor digitorum communis (Figure 3.22b) revealed a significant interaction between the two IVs ($ATS = 5.85, p < 0.05$) and a main effect of interaction technique ($ATS = 10.45, p < 0.01$) as well as posture ($ATS = 11.98, p < 0.001$). Investigation of the interaction showed that with M&K the mean frequency was higher when standing than sitting ($W = -132, p < 0.001$), and that the mean frequency was higher with M&K than W&N ($W = 122, p < 0.01$) when standing.

The analysis for right superior trapezius (Figure 3.22c) revealed no significant interaction between the two IVs and no significant main effect of posture. A significant main effect of interaction technique was

observed ($F(1, 64) = 4.82, p < 0.05$): mean frequency values were lower with M&K than W&N.

The analysis for right biceps brachii (Figure 3.22d) revealed no significant interaction between the two IVs and no significant main effect of posture. A significant main effect of interaction technique was observed ($F(1, 64) = 43.11, p < 0.001$): mean frequency values were lower with W&N than M&K.

3.5.7 EMG gradients

EMG gradients values were not normally distributed. EMG gradients data for right extensor digitorum communis was increased by a constant to become equal or greater than 1, in such a way that a square root transformation could be applied. Data recorded from other muscles could not be normalized.

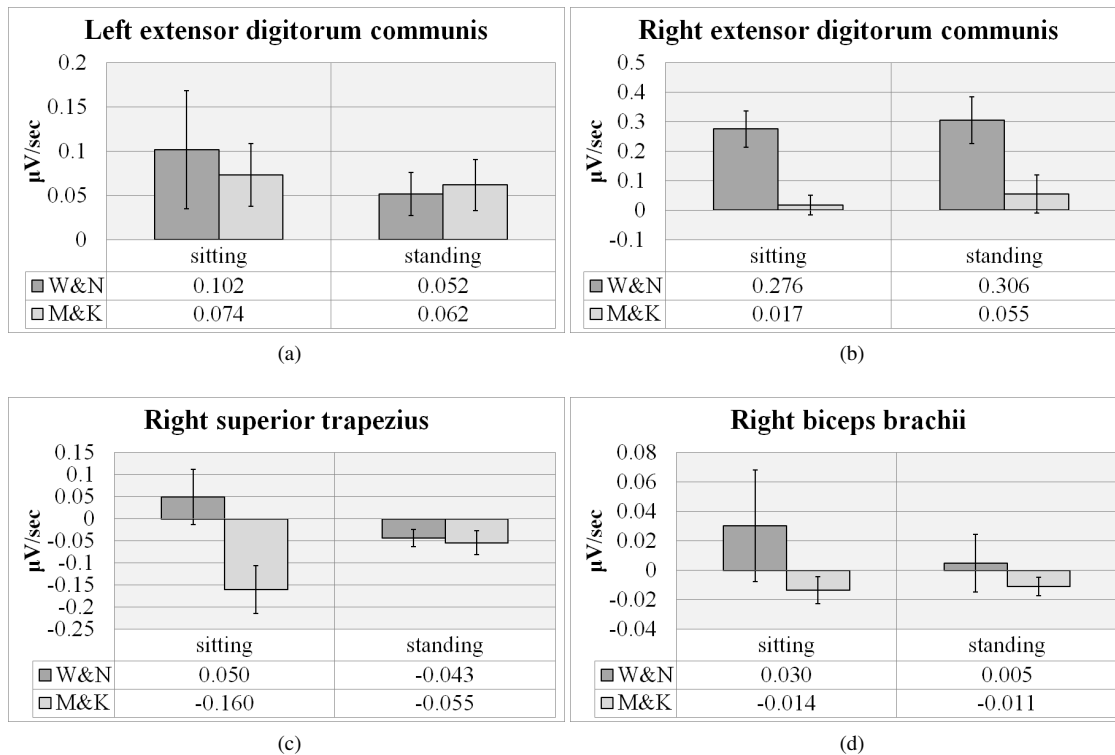


Figure 3.23: Change of EMG gradient data for the four muscles. Error bars indicate standard error of the mean.

ATS analysis of EMG gradient for left extensor digitorum communis (Figure 3.23a) revealed that neither interaction between the two IVs nor main effects IVs were significant. The analysis for right extensor digitorum communis (Figure 3.23b) showed no significant interaction between the two IVs and no significant main effect of posture. A significant main effect of interaction technique was found ($F(1, 60) = 14.15, p < 0.05$): the muscle activity showed a steeper increase over time with W&N than with M&K.

ATS analysis of EMG gradient for right superior trapezius (Figure 3.23c) revealed a significant interaction between the two IVs ($ATS = 4.08, p < 0.05$) and a main effect of interaction technique ($ATS = 18.13, p < 0.001$). Investigation of the interaction showed that EMG gradient was higher with W&N than M&K ($W = -125, p < 0.01$) when sitting.

ATS analysis of EMG gradient for right biceps brachii (Figure 3.23d) revealed that neither interaction between the two IVs, nor main effects were significant.

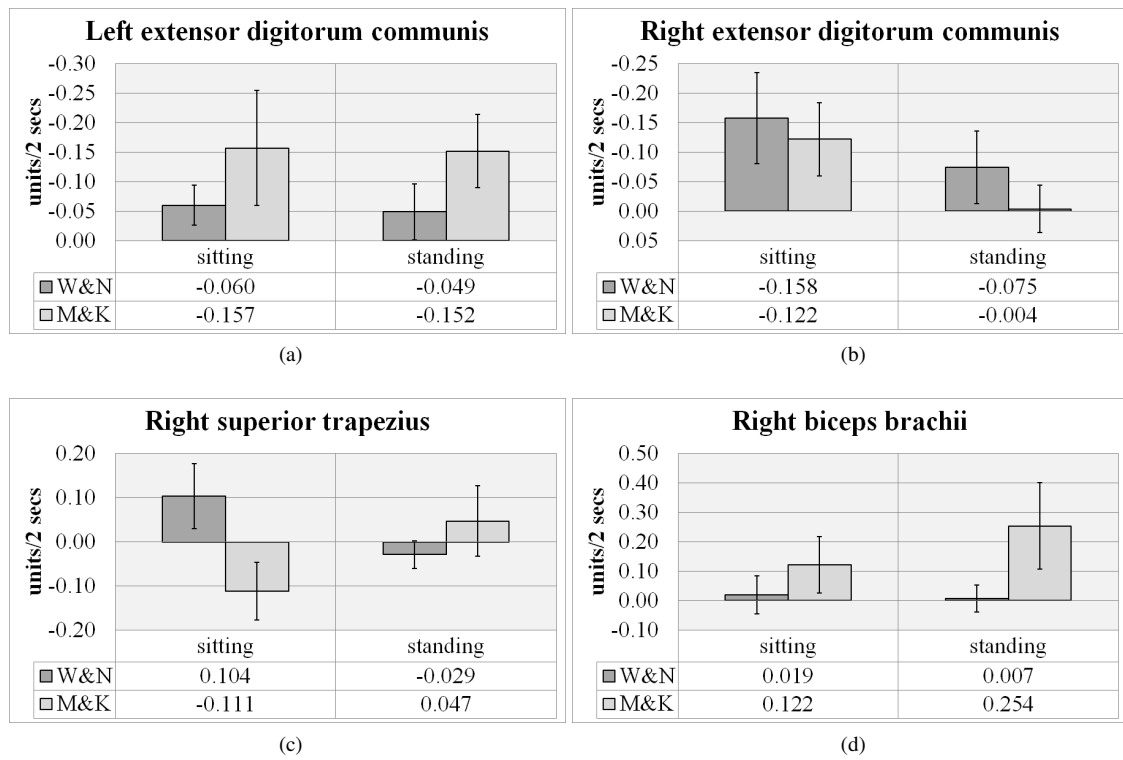


Figure 3.24: Linear regression slope of left and right extensor digitorum communis (charts with reversed scale), right superior trapezius and right biceps brachii. Error bars indicate standard error of the mean.

3.5.8 Linear regression slopes of the mean values of the EMG power spectrum mean frequency

Linear regression slope values were not normally distributed. Only left extensor digitorum communis slopes could be normalized by increasing values by a constant to become equal or greater than 1, in such a way that a square transformation could be applied. The analysis of linear regression slopes did not find significant differences for any analyzed muscles.

3.5.9 Ranked choice questionnaire

Ranked choice questionnaire data was analyzed with ATS. Mean values are shown in Figure 3.25.

ATS analysis of first and second item data (Figures 3.25a and 3.25b) revealed no significant interaction between the two IVs and no significant main effect of posture. Main effect of interaction technique was significant (ease: $ATS = 3.99$, $p < 0.05$; comfort: $ATS = 9.92$, $p < 0.05$): navigation was perceived as easier and more comfortable with M&K than W&N.

ATS analysis of third item data (Figure 3.25c) revealed that neither interaction between the two IVs nor main effect of interaction technique were significant. Main effect of posture was significant ($ATS = 3.93$, $p < 0.05$): participants found coarse manipulation easier when sitting than standing.

ATS analysis of fourth item data (Figure 3.25d) revealed no significant interaction between the two IVs while both main effects were significant (posture: $ATS = 4.00$, $p < 0.05$; interaction technique: $ATS = 19.04$, $p < 0.001$): participants found coarse manipulation more comfortable when sitting than standing, and more comfortable with M&K than W&N.

ATS analysis of fifth and sixth item data (Figures 3.25e and 3.25f) revealed that neither interaction between the two IVs, nor main effect of interaction technique were significant. Main effect of posture was significant (ease: $ATS = 3.90$, $p < 0.05$; comfort: $ATS = 4.83$, $p < 0.05$): fine manipulation was perceived as easier and more comfortable when sitting than standing.

ATS of seventh and eighth item data (Figures 3.25g and 3.25h) revealed neither significant interactions nor main effects.

3.6 Discussion of the Results

In this section, we discuss in detail the experimental results reported in Section 3.5.

3.6.1 User performance

As the analysis has shown (Section 3.5.1), it took participants more time to complete the task with W&N than M&K. This result confirms the findings presented in the literature (e.g., MacKenzie et al., 2001; Olsen Jr and Nielsen, 2001; Myers et al., 2002) but on a task (object arrangement) that to the best of our knowledge has not been considered before in the evaluation of LPS interaction techniques. Considering the detailed analysis of how participants spent time with W&N, it turns out that they navigated more, performed less coarse manipulation and more fine manipulation (which generally requires more time than coarse manipulation) with respect to M&K.

Differences in object manipulation time between the two interaction techniques can be better understood by looking at the way participants used the two devices for manipulation. To keep the on-screen cursor in the current position, the user does not need to apply any force to the mouse, while she needs to continuously hold the LPS interaction device in hand and point steadily at the screen, which results in more fatigue and muscle strain (see Section 3.6.2). Similarly, to move the cursor over a small distance, the mouse requires much less force than the LPS interaction device, and to perform fine manipulations with M&K, participants have to move their hands from the mouse to the keyboard arrows, and once they are done back again to the mouse.

This could explain why, with W&N, participants spent more time using fine manipulation controls, which do not require to keep the device pointed at the screen. Indeed, by observing the participants, we

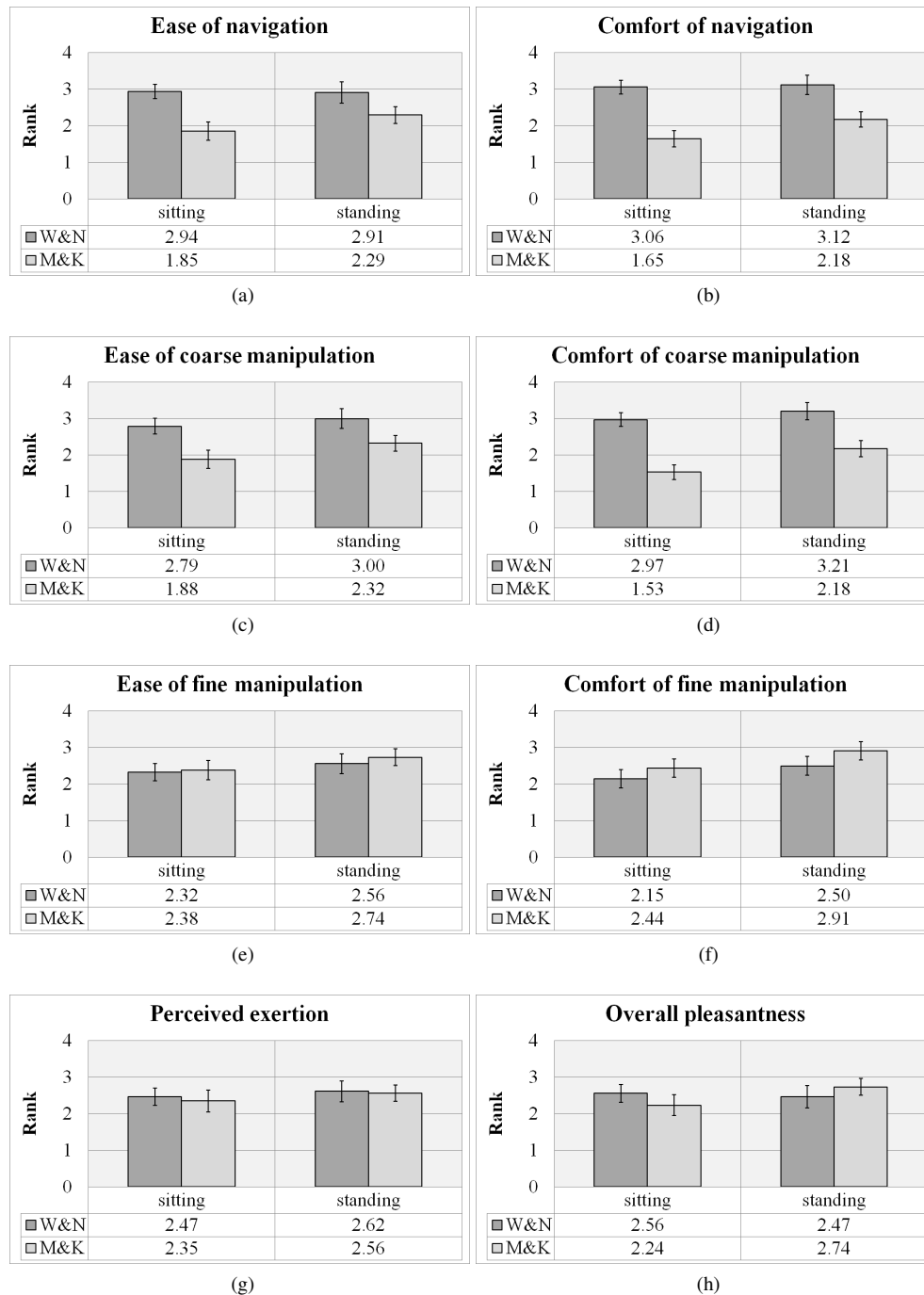


Figure 3.25: Mean values for the eight items of the questionnaire. Ranks range from 1 (best) to 4 (worst). Error bars indicate standard error of the mean.

noticed that, with the Wii Remote, they tended to make large changes with coarse manipulation to put the object roughly near the target, and then relying only on fine manipulation.

The difference in cursor latency, discussed in Section 3.3.1, might also have played a role, making users prefer fine manipulations (in which latency is reduced). Cavens et al. (2002) and Teather et al. (2009) pointed out that cursor latency affects negatively users' performance in Fitts and Tunnel tasks (Accot and Zhai, 1997). Moreover, the non-linear relationship between mouse and cursor speeds (the slower the mouse moves, the smaller is the distance travelled by the cursor) could have played a role: moving the cursor over small distances might have been easier with M&K because, with small mouse movements, the cursor can be controlled with a greater precision than W&N. As reported by Casiez et al. (2008), however, the effect of cursor acceleration on user performance is quite small.

3.6.2 Muscle activity

Left and right extensor digitorum communis The analysis showed that the mean activity of left and right extensor digitorum communis is affected by posture: it is higher when the task is performed in a standing rather than sitting position. A factor that contributes to explain this result is that using the chair with the armrests allows for a position that puts less strain on extensor digitorum communis muscles.

Results also showed that power spectrum mean frequency was lower with W&N than M&K, and was higher when standing than when sitting. Since lower values are an indication of a sustained muscle contraction and a signal of localized muscle fatigue (Merletti et al., 1990), these results indicate that muscle activity was more fragmented with M&K than W&N, giving participants more opportunities to recover from strain. This is also confirmed by total activity data, which showed that the accumulated activity of each muscle was greater with the LPS interaction device than M&K. We can hypothesize that participants applied a more constant force to fingers with W&N to keep the LPS interaction device constantly pointing at the screen. Similarly, results show that the standing position (which can be considered the most advantageous posture for LPS interaction with large displays because of the mobility afforded) requires a more constant muscle exertion with W&N than M&K.

A significantly steeper EMG gradient for the right extensor digitorum communis was also observed with W&N. Andreassi (2007) reviews various findings in the literature which correlate this kind of result with user involvement in the task. Thus, the use of the LPS interaction device might require a greater level of user involvement to successfully complete the task with respect to M&K, perhaps due to the greater muscle control required to keep the device pointing at the screen during the task.

Right biceps brachii The analysis of mean and total biceps activity revealed that W&N elicits a greater muscle activity with respect to M&K. This is not surprising, since the LPS interaction device requires to point at the screen keeping the forearm lifted and the biceps under load, thus causing strain. Even if the chair used during the experiment was provided with armrests, the analysis revealed no significant difference between EMG values recorded when sitting or standing. As we observed during user evaluation, users generally leaned their right elbow on the armrest, but not their forearm. power spectrum mean frequency data confirms that the LPS interaction device required a sustained muscle contraction, thus causing more muscle fatigue.

The greater biceps activity might have consequences on the motor system if used for prolonged time: Sommerich et al. (2006) reports various studies which correlate sustained muscle contraction in the upper extremities with growing risks of cumulative disorders.

Right superior trapezius The analysis of trapezius data shows that its mean activity was greater with M&K than W&N, and that the mean frequency of the power spectrum was higher with W&N than M&K. We thus hypothesize that the shoulder muscle was generally less strained by the use of the LPS interaction device, because the force needed to handle it was localized more in the biceps brachii. While mean and total muscle activity with M&K was greater when sitting rather than standing, the opposite was observed for W&N, probably thanks to the use of the chair armrests. More interestingly, the total muscle activity with the LPS interaction device was greater in the standing than sitting posture, although, as we pointed out before, the standing posture could be considered more advantageous for LPSI with large displays.

Finally, EMG gradient data in the sitting posture was significantly higher with W&N than M&K, suggesting again that the use of the LPS interaction device requires a greater level of involvement.

The shoulder complex provides the greatest range of motion of all the body joints, at the cost of reduced joint stability and potential for entrapment of various soft tissues when the arm is elevated or loaded (Sommerich et al., 2006): various structures under the shoulder complex (the *brachial plexus* muscle, the *subclavian artery* and shoulder nerves) can be compressed by muscles or bones when the humerus is elevated or when the shoulder is loaded indirectly, i.e., when holding a load in the hand. In this context, the superior trapezius, together with other scapular muscles, elevates the shoulder to position and move the arm in space for the purpose of hand function (Clarkson, 2000). This suggests that superior trapezius and general shoulder activity are strongly related, as pointed out by Aarås (1994). The trapezius can become strained when the arm is held in an elevated position for long periods (Scheumann, 2007). Therefore, the EMG data we collected, especially considering the differences caused by posture, should call attention to the effects on the shoulder caused by the use of LPS interaction devices for extended periods of time, even if their weight is relatively small (in our study, it was about 140 g). Rozmaryn (2005) reports that if a load is held in the hand, the load moments at the elbow and the shoulder can become large relative to the flexor tendon moments required at both joints. Thus, even small loads cannot be supported for sustained periods (Rozmaryn, 2005), especially if the arm or forearm is elevated and pushed forward.

3.6.3 Other physiological measurements

The analysis showed that posture greatly affected the circulatory system, as expected (Jones et al., 2003). Moreover, the significant difference in peripheral temperature recorded with the two interaction techniques can be explained by noting that Nunchuck and keyboard have to be handled differently, thus eliciting differences in peripheral blood circulation at the left hand measurement site. More generally, participants' left hand was cooler during the evaluation with respect to the baseline (see Figure 3.17) because their forearm were generally kept elevated and stationary, causing poor blood circulation in peripheral vessels.

Overall, circulatory, respiratory and temperature measurements did not provide highlight important differences in mental stress between the two interaction techniques: the availability of fine manipulation probably mitigated the mental stress caused by the less precise interaction with objects offered by coarse manipulation, and users did take advantage of fine manipulation.

3.6.4 Subjective preferences

Participants perceived W&N as more difficult and less comfortable than M&K during coarse manipulation. This is consistent with the observations about muscle activity reported earlier. The difference in perceived ease of use could also be explained by the fact that all participants had at least basic experience and were familiar with M&K (as reported in Section 3.4.1), while less participants had used W&N before the evaluation.

Participants' perception of navigation is also characterized by more difficulty and less comfort with W&N than M&K. We observed that, while the Nunchuck did not require to point at the screen, participants tended to keep the left forearm elevated at the same height of the right arm during task execution, thus possibly reducing comfort. Surprisingly, even though navigation with W&N was perceived as more difficult than with M&K, some participants' informal comments reported that they appreciated the use of the Nunchuck joystick to move inside the VE. A few participants who play more frequently video games on PC or the Wii reported that they found initially unnatural to use left and right arrows on the keyboard and left and right joystick movements on the Nunchuck to change the viewpoint orientation instead of strafing (see Section 3.3.2), but they quickly got used to it.

No clear preference between the two interaction techniques emerged with fine manipulation, probably because fine manipulation is carried out in both cases by pressing buttons (without the need to point at the screen), and objects move, rotate and scale at the same speed.

Generally, participants preferred the sitting posture to carry out the tasks, and perceived sitting as more comfortable, probably thanks to the fact that they could use the armrests or the desk to lean their forearms. Furthermore, the standing posture is generally more fatiguing: see, for example, the postural effects on the circulatory system, reported in the previous section.

Differences in perceived fatigue and pleasantness with the two interaction techniques were consistent with surface EMG analysis, but they were small and did not reach statistical significance.

4

Physiology–Based Evaluation of Video Game User Experience

In Chapter 3, we investigated the use of physiological signals, focusing in particular on EMG, in the evaluation of the bodily responses elicited by the use of a novel interaction device. In this chapter, we go one step further, and shift our focus on the relations between users' bodily responses and their emotions, on the basis of the literature reviewed in Section 1.3. In particular, in this chapter we explore the effectiveness of physiological signals for the assessment of users' emotions afforded by a violent mobile video game (Chittaro and Sioni, 2012b). Violent video games are often associated to negative effects such as desensitization to violence. However, while aggression can concern any living being, experiments in the literature have especially focused on games that require the player to aggress human (or anthropomorphic) beings. To extend the investigation of violent video games, this chapter considers a video game genre (*Whac-a-Mole*) in which the victims of aggression belong to non-human animal species. The study investigates UX aspects (in terms of players' emotions) as well as desensitization to violence aspects of a Whac-a-Mole game and a non-violent version of the same game, using Affect Grids and physiological measures (Facial EMG, SCL, HR, and BVPA). To obtain a high level of control on confounding factors, the modified game for the non-violent condition of the study replaces only the violent content of the original game with non-violent content, leaving all other game features constant. Well-known findings about desensitization to violence in violent video games were not found in this study, and player's affect results also suggest that the violent element of the Whac-a-Mole game cannot be straightforwardly replaced by a non-violent one without possibly weakening the User Experience.

4.1 Introduction

The classic *Whac-a-Mole* arcade game is the first example of a popular genre (*Whac-a-Mole* games) in which a specific violent element (i.e., aggression towards non-human animal species) seems to be central to the User Experience (UX). ISO 9241-210 defines UX as “a person's perceptions and responses that result from the use or anticipated use of a product, system or service.” In the Whac-a-Mole arcade game, plastic moles pop up from the game cabinet, and players have to hit each mole using a soft mallet to force it back into the cabinet. The more moles the player hits, the higher the score. Subsequent Whac-a-Mole games employed various non-human animal species besides moles. As video game technology progressed over the years, this game genre moved to computers and, more recently, to handheld consoles (e.g., Nintendo DS) and mobile phones (e.g., Apple iPhone), which are well suited for quick, on-the-go gaming sessions afforded by this genre.

Several studies in the literature have contrasted violent and non-violent video games, concluding that playing violent video games results in negative effects such as an increase in negatively-valenced high-arousal emotions (e.g., Ravaja et al., 2008) such as anxiety (e.g., Anderson and Ford, 1986), an increase of aggressive feelings, cognition and behaviors (e.g., Anderson and Dill, 2000; Anderson et al., 2004), and desensitization to violence (e.g., Bartholow et al., 2006; Carnagey et al., 2007; Engelhardt et al., 2011).

While aggression is defined as behavior directed towards the goal of harming other living beings (Baron et al., 2009), the experiments in the literature have mainly focused on games that require the player to

aggress human (or at least anthropomorphic) beings. The classic *Whac-a-Mole* arcade game is instead the first example of a popular genre (Whac-a-Mole games) in which a specific violent element (i.e., aggression towards non-human animal species) seems to be central to UX. In the Whac-a-Mole arcade game, plastic moles pop up from the game cabinet, and players have to hit each mole using a soft mallet to force it back into the cabinet. The more moles the player hits, the higher the score.

In this chapter, we present an experimental evaluation that exploits many of the users' physiological signals described in Section 1.3 to compare the UX of two Whac-a-Mole games. The first of the two video games is a traditional Whac-a-Mole game that requires the player to kill insects, and the second is a modified version of the first game in which the violent element is replaced by a non-violent one.

An issue that characterizes experimental comparisons of violent and non-violent video games in the literature concerns confounding variables (Adachi and Willoughby, 2011). The violent and non-violent game conditions are typically based on completely different video games (e.g., Arriaga et al., 2006; Carnagey et al., 2007; Sestir and Bartholow, 2010), such as a FPSs in the violent condition vs. a puzzle game in the non-violent condition. To address this issue, some studies try to match violent and non-violent video games on additional dimensions other than violent content (e.g., enjoyment and arousal) or try to employ multiple video games for both the violent and non-violent conditions to mitigate the stimulus sampling problem (Wells and Windschitl, 1999). Unfortunately, the violent and non-violent conditions can differ in a very large number of aspects (e.g., sounds, music, pace, difficulty, environment in which the game takes place, structure of the game levels, nature and timing of game events, sequence and type of actions required to the player, motor actions on the game controls, etc.). With so many game characteristics not held constant between conditions, one cannot rule out that unaccounted for variables—instead of violent content—can be responsible for the differences obtained, as discussed in depth by McMahan et al. (2011). To face this methodological issue, we changed the source code of the studied violent video game, producing a non-violent version of the same game that differs only in the nature (violent or non-violent) of the player's action.

The goal of the study presented in this chapter is to assess (i) the possible effects of the violent element on users' emotional state, and (ii) the possibility of effectively replacing their violent element with a non-violent one. Moreover, we study possible desensitization to violence effects after playing both games.

The chapter is organized as follows. In Section 4.2, we briefly review the literature about the use of physiology for evaluating video game UX, as well as about negative effects of violent video games. The two versions of the game employed in our study are described in Section 4.3. Then, Sections 4.4 and 4.5 describe in detail the experiment and its results, while Section 4.6 critically discusses the results.

4.2 Related Work

4.2.1 Physiological measurements and video game UX

Correlations between facial EMG measures and positively or negatively-valenced emotion, as reported in Section 1.3.2, are well known in the psychophysiology literature, and are commonly observed also in the UX literature. For example, Hazlett (2006) observed that zygomaticus major activity was greater during positive events (e.g., overtaking opponents' cars or winning a race) than negative events (e.g., being overtaken or hit by another car), while corrugator supercilii activity was greater during negative events, consistently with the findings summarized in the Section 1.3.2. When considering violent video games such as FPSs, physiological data can highlight more complex reactions: Ravaja et al. (2008) observed that killing human opponents in a commercial FPS elicited a fall of zygomaticus major activity as well as an increased EDA, suggesting high-arousal negative feelings rather than the pleasant feelings that can result from winning and success. On the contrary, the wounding and death of the player's own character elicited an increase of zygomaticus major activity and EDA as well as a decrease of corrugator supercilii activity, which could be possibly explained by a transient relief from (stressful) engagement.

Nacke and Lindley (2009) tried to characterize terms used in the game UX literature—such as immersion, boredom, fun, and *flow* (Csíkszentmihályi, 1990)—in terms of physiological measurements of player's arousal and valence. To this purpose, they designed three levels of a FPS characterized respectively by boredom (weak enemies, repetitive and overly simplistic environment, no surprises), immersion (complex and

appealing environment, strong as well as weak enemies) and flow (single game mechanic, level of increasing difficulty). Participants played the three levels while facial EMG and SCL data were recorded. Results show that challenging situations (which mainly characterized the flow level) elicited high-arousal positive affect in players.

Drachen et al. (2010) found instead some correlations between physiological measures (HR and EDA) and gameplay experience components measured with the *Game Experience Questionnaire* (GEQ) (IJsselstein et al., 2008), which assesses 7 dimensions of the video game UX (flow, challenge, competence, tension, negative affect, positive affect, sensory and imaginative immersion). In particular, authors found (i) a positive correlation between HR and tension, HR and negative affect, EDA and negative affect; (ii) a negative correlation between HR and several GEQ dimensions (competence, immersion, flow, challenge, and positive affect). Other studies addressed the same topic by correlating physiological data and subjective emotion assessment. In (Mandryk and Atkins, 2007), cardiovascular and respiration data as well as EMG (measured on the jaw) and EDA were first recorded while participants were playing a commercial hockey video game against the computer or a friend. After the gaming session, a subjective questionnaire was proposed to rate how boring, challenging, easy, engaging, exciting, frustrating and fun each of the two conditions was. Authors observed that playing the game with a friend elicited greater EDA values and that physiological results were correlated with questionnaire scores: EDA was positively related to fun and negatively related to frustration; jaw EMG activity was positively related to boredom as well challenge, but was negatively correlated with ease.

4.2.2 Effects of violent video games

4.2.3 Effects of violent video games on aggressive feelings, thoughts and behaviors

Many studies in the literature have associated violent content in video games to aggressive feelings, thoughts and behaviors (Anderson et al., 2010; Boyle et al., 2011). In most of these studies, participants typically played a violent or a non-violent game, and the aggression elicited by playing the game was later measured through subjective questionnaires or specific post-play tasks. In some cases, physiological measurements were also employed, as we summarize in the following.

In (Anderson and Dill, 2000), authors recruited participants who, during mass testing questionnaire sessions, scored low or high in the *Caprara Irritability Scale* (CIS) (Caprara et al., 1985). The participants were asked to play a violent game (a FPS) or a non-violent one (a graphic adventure), letting them believe that they were competing against a second player over a network. After the gaming session, their aggressive behavior was measured using a version of the *Competitive Reaction Time Task* (CRTT) (Taylor, 1967). In this post-play task, participants have to press a button as quickly as possible in response to an audio stimulus and are led to believe that they should react faster than a competitor. Out of 25 time trials, the CRTT makes the real player win 13 times and the simulated player 12 times. At the end of each time trial, the loser receives an aversive punishment (e.g., loud noise), the intensity of which is supposedly set by the opponent. Prior to each trial, each participant sets the punishment level (e.g., the intensity and duration of the noise) that will be delivered to the opponent if the participant wins the trial. The punishment level is used as a measure of aggressive behavior. Participants who obtained higher scores in the CIS as well as participants who played the violent video game showed greater aggressiveness in the CRTT. Two physiological measurements (heart rate, blood pressure) were also analyzed, but did not provide significant results.

Anderson et al. (2004) obtained similar relations between violent video game exposure and aggressiveness in subsequent experiments concerning five violent video games (three FPSs, a combat driving game, a fighting game) and five non-violent video games (a pinball game, a strategy game, a puzzle game, a racing game, and a graphic adventure). They employed the tests and physiological measurements previously mentioned as well as the *Word Completion Task* (WCT) (Anderson et al., 2003). This task provides a list of 98 words with letters missing in strategic positions (e.g., *explo_e*), in such a way that half of these words can be completed to form a non-aggressive word (*explore*) or an aggressive one (*explode*), while the other half do not lend themselves to aggressive completions. Participants were given 3 minutes to complete as much of the WCT as they could. An accessibility of aggressive thoughts score was calculated for each participant by dividing the number of aggressive word completions by the total number of completions.

Authors observed that violent video games in general increased the accessibility of aggressive thoughts, and that trait hostility and trait aggression positively correlated with aggressive behaviors carried out during the experimental tasks.

Sestir and Bartholow (2010) analyzed the effects of violent and non-violent video games on aggressive thoughts and behaviors, but also on prosocial outcomes. They carried out three experiments in which participants had to play two FPSs as violent games or two puzzle games as non-violent games, and their aggression-related feelings were subsequently assessed using the *Interpersonal Reactivity Index* (IRI) (Davis, 1983), the CIS, the *Aggression Questionnaire* (AQ) (Buss and Perry, 1992), an aggressive affect scale developed by Anderson et al. (1995), the WCT, the CRTT, and the *Story Completion Task* (SCT) (Rule et al., 1987) which assesses participants' reactions to potential conflict situations by requiring them to complete a series of story stems. To evaluate the temporal stability of acute game exposure effects, in the first two experiments the aggression-related outcomes were assessed either immediately or after a brief delay. Reported differences indicate that violent game exposure increases aggression, while non-violent video game exposure decreases aggressive thoughts and feelings as well as aggressive behavior. In the analysis of prosocial outcomes, exposure to non-violent games with no explicit prosocial content increased prosocial thoughts, relative to both violent game exposure and, on some measures, a no-game control condition.

Anderson and Dill (2000) hypothesized that violent video games and aggressive behavior are related because users tend to learn aggression-related scripts by playing violent video games and then act them in real-life situations. Bushman and Anderson (2002) corroborated this hypothesis by using the SCT: in their experiment, the number of aggressive responses in the fictional stories was greater for participants who previously played a violent video game.

Anderson and Carnagey (2009) examined the impact of excessive violence in sports video games in three experiments. Participants' exposure to violent games and in particular to violent sports video games were assessed with a version of the *Video Game Questionnaire* (VGQ) (Anderson and Dill, 2000), in which the exposure to violent games is obtained by combining the time spent to play five of the participant's favorite games and a subjective rating of their violence content. Participants' aggressive feelings were evaluated using a subset of the AQ. After playing a randomly assigned violent or non-violent sports game, participants rated it using a video game evaluation questionnaire (Anderson and Dill, 2000) and completed the (i) *Word Pronunciation Task* (WPT) (Anderson et al., 2003), in which the reaction time to pronounce aggressive and non-aggressive words is measured; (ii) the *State Hostility Scale* (SHS) (Anderson et al., 1995); (iii) the *Attitude Towards Violence in Sports Questionnaire* (ATVS), originally created for the experiment and (iv) the CRTT. During these three phases, cardiovascular measurements were recorded (blood pressure and pulse). Results show that violent sport games increase aggressive affect, cognition and behavior, as well as attitude towards violence in sports; no differences were observed with regards to the effects of violent and non-violent sport games on the cardiovascular measures. Unlike most studies in the literature, violent and non-violent games were similarly competitive in this study: this allows to exclude a difference in game competitiveness as a possible explanation for the obtained results. However, Adachi and Willoughby (2011) point out that it does not allow one to conclude that the results are uniquely related to the aggressive elements of the considered games, because the games were not equated on other important dimensions such as difficulty and pace of action.

Instead of comparing violent and non-violent games, Barlett et al. (2008) concentrated on a specific aspect of the portrayal of violence that is used in some video games. They carried out an experiment in which participants played the same violent fighting game with four different levels in the amount of blood displayed on the screen (maximum, medium, low and off), while their HR was recorded to assess physiological arousal. Before the gaming session, participants' trait aggression and aggressive feelings and thoughts were assessed with the AQ and the SHS, while aggressive behavior was calculated as the ratio of total weapon time (i.e., how long the participant used a weapon during the gaming session) to total gaming time. Results show that participants in the maximum and medium blood conditions had a significant increase in hostility and arousal with respect to participants in the other two conditions. Furthermore, those in the maximum and medium blood conditions used character's weapon significantly more than the other participants.

However, it must be noted that the literature is not unanimous about the impact of violent video games on aggression. In his meta-analytic review, Ferguson (2007) reasons that the research work on violent video games is subject to a publication bias, i.e., the studies showing significant effects of violent content

on aggression appear to be over-represented in the literature (Boyle et al., 2011). Once corrected for publication bias, Ferguson claims that studies of video game violence could not provide support for the hypothesis that violent video game playing is associated with higher aggression (Ferguson, 2007).

4.2.4 Effects of violent video games on desensitization to violence

Some studies in the literature have focused on possible negative effects that playing violent video games could have in terms of desensitization to violence, using physiological measurements. In (Bartholow et al., 2006), EEG signals were recorded from participants while they were watching a series of images taken from the *International Affective Picture System* (IAPS) (Lang et al., 1997). Afterwards, participants performed a variant of the CRTT. Smaller amplitude values of P300 signals (which measure brain responses to relevant task-related stimuli) were observed in violent video game players and participants with aggressive traits, assessed using the VGQ, the CIS and the AQ.

A study by Engelhardt et al. (2011), based on P300 measures, has linked desensitization to violence with increased aggression. Participants were randomly assigned to a gaming session with violent (three FPS and a third-person action game) or non-violent games (two platform games and two sports games). After the play session, the P300 component of the event-related brain potential was recorded from participants while they were watching neutral and violent images taken from the IAPS used as stimuli for the P300 responses, to assess desensitization effects. Finally, participants completed a CRTT, to measure the level of aggressiveness. Authors report that for participants whose prior exposure to violent video games was low, playing violent video games caused a reduction in the brain's response to violent images, and an increase in aggressive behaviors. This recent finding links violence desensitization with findings on increased aggression reported in the previous section, providing evidence that desensitization to violence can account for changes in aggression.

The study described in (Carnagey et al., 2007) used HR and EDA signals instead of EEG. Participants played a randomly assigned violent or non-violent video game, chosen among four violent games (a FPS, a fighting game, a violent racing game and an action game) and four non-violent games (a puzzle game, a strategy game, a platform game and a pinball game). Afterwards, they watched real-life violence episodes (courtroom outbursts, police confrontations, shootings and prison fights) recorded on video. Participants who previously played a violent video game had lower HR and EDA while viewing the violent video footages, suggesting desensitization to violence.

4.3 The Considered Video Games

The experiment employs one of the recent Whac-a-Mole video games for the iPhone, in which the aggressed animals are insects (wood ants) that the player has to squash using her fingers. To choose the game, we initially analyzed games rated for frequent/intense violence in the Apple App Store. The company that produces one of them consented to give us access to the game programming code for the purpose of this study. The considered game shows a realistic 3D model of a chair at the center of the screen (see Figure 4.1).

Players can rotate the chair by performing swiping gestures on the screen or by touching two rotation buttons at the bottom of the screen. By swiping the screen from right to left (resp. left to right) or by touching the left (resp. right) rotation button, the chair rotates clockwise (resp. counterclockwise) by 90°. A vertical bar on the left side of the screen indicates the status of the chair: at the beginning of a level, the bar is full and green, indicating that the chair is in perfect conditions. Wood ants appear randomly on the chair, walk on it for a few instants and then bite it, producing a munching sound and leaving wood shavings on the floor (Figure 4.1a). The more the chair is damaged, the more the left bar drains and turns gradually to red (Figure 4.1a), pulsating more and more frequently. When the left bar is empty, the chair breaks down and the game is over (Figure 4.2).

To successfully complete a level, players have to squash ants by touching them with fingers on the screen. The blue bar placed on the right side of the screen grows when insects are killed. To successfully complete a game level, the player has to entirely fill the blue bar before the ants damage too much the chair leading to its break down. Every time an ant is killed, the game plays a crushing sound, and the

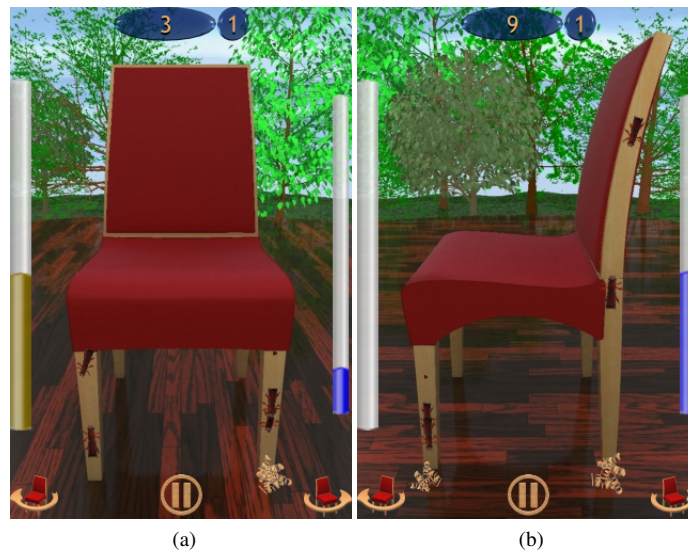


Figure 4.1: Screenshots of the violent Whac-a-Mole game: (a) the wood ants are walking on the chair; (b) two ants are squashed. Note the wood shavings on the floor.

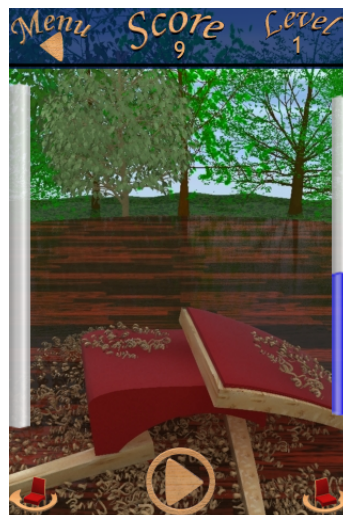


Figure 4.2: *Game Over* screen with the broken chair.

image of the insect changes to realistically represent a squashed ant, which disappears after a few seconds (Figure 4.1b). At the top of the screen, two blue ovals show respectively the number of ants killed during the gaming session (i.e., the total score) and the current game level. The higher the game level, the more frequently ants appear on the chair. To create a non-violent version of the studied game, we changed the game programming code to remove the violent element. The change we made was to replace ants with abstract geometric shapes of the same size, eliminating the need to do harm to virtual living beings in order to progress in the game. More specifically: (i) live ants were replaced by purple rectangles (Figure 4.3a), (ii) the biting ant animation was replaced by a quick pulsating rectangle animation, and (iii) squashed ants were replaced by fading animations of a gray rectangle (Figure 4.3b). Any other detail of the game, including sound, remains identical.



Figure 4.3: Screenshots of the non-violent Whac-a-Mole game: (a) two purple rectangles appear; (b) the user has touched the two rectangles and a new rectangle is appearing.

4.4 Experimental Evaluation

Our study follows a between-subjects design with *game* (*violent* and *non-violent*) as the independent variable, and concentrates on measuring possible differences in affect during gameplay as well as desensitization to violence after gameplay. In this section, we describe in detail the experiment.

4.4.1 Participants

The evaluation involved a sample of 42 participants (32 M, 10 F) recruited among graduate and undergraduate students with various educational backgrounds (computer science, literature and philosophy, physiotherapy, natural sciences). Age ranged from 20 to 39 ($M = 25.1$, $SD = 3.8$). We employed the VGQ (Anderson and Dill, 2000) to assess individual differences in the time participants spend on violent video games: the version of the questionnaire we used asks participants to indicate no more than five of their most preferred video games, played during the last two years, and rating them in the 1–5 range with respect to how often they played each game and its perceived degree of violence. A VGQ score, in the 0–25 range, is calculated as the average of the product, for each game, of the time spent playing it and its degree of violence, subjectively assessed by the participant. To assess individual differences in physical aggressiveness and anger traits of participants, we employed the corresponding sets of items (16 items in total) from the AQ (Buss and Perry, 1992) on a five-level Likert scale. The AQ scores can range from 16 to 80. Finally, we assessed participants' attitudes towards insects, using two items (“I find insects disgusting” and “It can happen that

I intentionally squash an insect”) on a five-level Likert scale (1 = “Not at all typical of me”, 5 = “Very typical of me”). The means for the above mentioned individual differences were very similar in the two groups. More specifically, mean AQ score was 33.19 ($SD = 7.02$) in the non-violent and 33.10 ($SD = 6.65$) in the violent game condition, while mean VGQ score was 5.82 ($SD = 4.74$) in the non-violent and 5.41 ($SD = 4.25$) in the violent game condition. Means for the first item concerning attitudes towards insects in the two groups were respectively 2.71 ($SD = 1.45$) and 2.57 ($SD = 1.29$), while for the second item they were 2.95 ($SD = 1.47$) and 2.71 ($SD = 1.55$). For each of the four pair of means, a t-test indicated that the reported small differences between the two groups of participants were not statistically significant.

4.4.2 Materials

The games were run on an Apple iPhone 3GS equipped with a 3.5”, 320×480 pixels screen. During the gaming session, the device was lying on a table, and participants were seated on a chair. To record participants’ physiological data, we employed six sensors: four EMG sensors for facial EMG, coupled with disposable electrodes; a PPG for measuring BVP, recorded on the distal phalanx of the middle finger; two electrodes for EDA, placed on the intermediate phalanges of the index and the ring fingers (Figure 4.4).

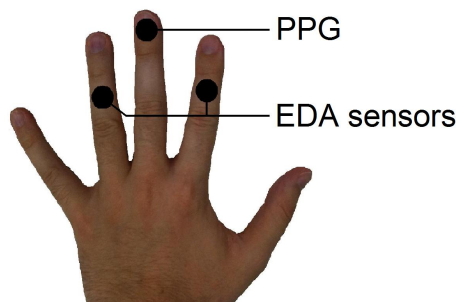


Figure 4.4: Position of PPG and EDA sensors during the experiment.

The signals were recorded with a Thought Technology Procomp Infiniti encoder and stored on a PC. A second computer was used to show participants two video footages taken from movies, respectively showing: (i) a man intentionally squashing ants with increasing anger, (ii) a man savagely beating another man, even after the latter falls down bleeding.

As in the other studies summarized in Section 4.2.1, we used facial EMG data to determine the elicited emotional valence, and SCL, HR and BPPA measurements to assess the arousal level. For subjective emotion assessment, we employed *Affect Grids* (AGs, Figure 6) (Russell et al., 1989), which are also based on emotional valence and arousal.

In AGs, participants choose one of the 81 grid cells to indicate how they feel. For example, if a participant is experiencing pleasant feelings, (s)he can mark a cell on the right half of the grid; if a participant is depressed, (s)he can mark a cell near the bottom-left corner. From the position of the marked cell, two values in the 1–9 range are obtained: perceived level of arousal (arousal–sleepiness score, or *A* score) and pleasantness (pleasure–displeasure score, or *P* score).

4.4.3 Procedure

Participants were informed that the goal of the study was to evaluate the UX of a video game, and consented to have physiological signals recorded during the experiment. They were also clearly informed that all the experimental data was going to be collected and analyzed anonymously for research purposes, and asked if they consented to participate in the experiment. Then, they filled the previously described questionnaires. On the basis of the results of the questionnaires, participants were assigned to one of the two conditions to balance gender, time spent playing violent video games, physical aggressiveness and anger trait, and attitude towards insects. Participants were seated and the skin of their forehead and cheeks was cleaned using a

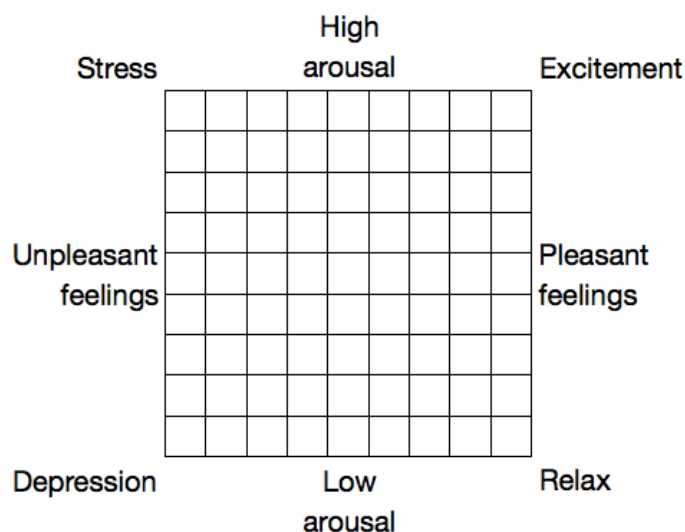


Figure 4.5: The AG template used in our study.

pad of cotton wool and alcohol, then the EMG sensors were applied following the electrodes placement recommended by Fridlund and Cacioppo (1986), depicted in Figure 4.6.

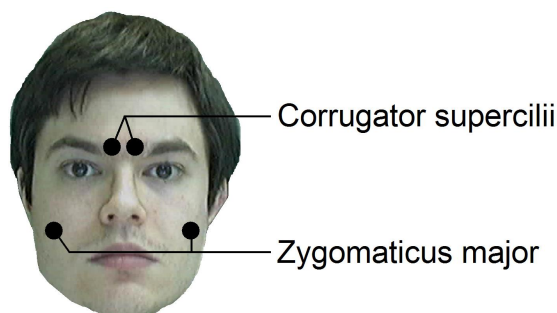


Figure 4.6: Location of left and right CS and ZM.

PPG and EDA electrodes were applied to the fingers (middle finger, index finger and ring finger respectively) of the non-dominant hand. Once the sensors were in place, participants were asked to relax for about three minutes to record the baseline for the physiological signals, i.e. the signals' values that can be observed when participants are in a resting state. When analyzing data, baseline values have to be subtracted from the data recorded during the experimental conditions, to separate the physiological responses to experimental stimuli from the intrinsic physiological differences among participants. During baseline recording, we asked participants to keep the non-dominant hand on the table or on the lap as steady as possible to reduce possible recording artifacts caused by arm movement. During baseline recording, a video with relaxing images and music was shown in a dim light. Participants could close their eyes and only listen to the music if they preferred. After baseline recording, participants filled a pre-play AG indicating the emotion felt at that moment, and then played the assigned game for 10 minutes with their dominant hand, while they sat comfortably.

We presented the games to participants as follows:

- Violent game: “In this game, ants move on a chair. When they bite, they chop away wood from the chair. You have to squash them with your fingers to prevent the chair from breaking down”;
- Non-violent game: “In this game, purple rectangles move on a chair. When they flash, a piece of

the chair falls off. You have to touch them with your fingers to make them disappear and prevent the chair from breaking down”.

At the end of the gaming session, participants filled a post–play AG, indicating again the emotions felt at that moment. Then, they watched the two violent videos described in the Section 4.4.2. Before each video, participants were asked to relax for a minute, so that physiological parameters could revert to a rest state. After seeing the two videos, participants were allowed to remove all the sensors and were asked to fill two AGs, indicating the emotions felt while watching each of the two videos. Finally, participants were debriefed and thanked for their participation.

4.4.4 Measures

For each participant, we calculated the following physiological measures in the gameplay phase and in each of the two video watching sessions:

- *Mean activity of left ZM, mean activity of right ZM, mean activity of left CS and mean activity of right CS*: mean values of the root mean square (RMS) transformation of the four facial EMG signals, averaged over one second epochs. A notch filter (band–stop filter), centered on the 50 Hz frequency, was applied to remove typical AC interference caused by electronic devices;
- *Mean HR and mean BVPA*: to obtain the two mean values, signal was averaged over five second epochs to reduce artifacts caused, for example, by hand movements. Epochs are here longer than those used for muscle activity, because HR and BVPA are subject to slower variations;
- *Mean SCL*: the mean value of SCL measured during the task.

The subjective participants’ reports acquired with the four AGs were used to obtain the following measures for each participant:

- *P score delta and A score delta*: the difference in *P* score, and the difference in *A* score, between the post–play and pre–play AGs. This was calculated to assess if the participant perceived a change in affect after playing the assigned game;
- *Video watching AG scores*: the *P* score and *A* score reported by the participant, for each of the two videos.

4.4.5 Possible outcomes

The findings on violent video games reported in the literature and summarized in Section 4.2.2 would suggest that desensitization should be observed in players of the violent Whac-a-Mole game, who should show a reduction in physiological responses (as well as subjectively reported arousal) during violent video watching with respect to players of the non–violent version of the game. Moreover, since studies in the literature have shown that carrying out aggressive actions in video games elicits anxiety in players (Anderson and Ford, 1986) and, more generally, negatively–valenced high arousal affect (Ravaja et al., 2008), the violent element of the studied Whac-a-Mole game could have similar negative effects on UX, resulting in increased CS activity (i.e., an increase in intensity of negatively–valenced emotions) as well as decreased ZM activity (i.e., a decrease in intensity of positively–valenced emotions) and increased EDA (i.e., an increase in arousal) during gameplay with respect to the non–violent version of the game.

However, we should take into account that, unlike the studies on violent video games in the literature in which aggression concerns human or anthropomorphic animals, our study concerns a non–human and non–anthropomorphic animal species, which is a considerable difference that can change affective reactions of players. Indeed, aggressive behaviors towards animal species such as insects are much more socially tolerated than aggression towards humans (Martens et al., 2007). It is also more difficult to empathize with such species. Therefore, the outcomes of the study might not replicate those of the current violent video game literature.

For the reasons above, our study is exploratory in nature and is aimed at starting an investigation of possible analogies and differences between the effects of Whac-a-Mole video games and those reported for other violent game genres in the literature, as well as the possibility of effectively replacing the violent element in Whac-a-Mole games with a non-violent one.

4.5 Results

For conciseness, in the following we will refer to the violent game condition as *VG* and to the non-violent game condition as *NVG*. Before proceeding with statistical analysis, baseline values were subtracted from the physiological means to remove individual differences. *Interquartile Range* (IQR, i.e., the difference between first and third quartiles) was employed to find outliers among the resulting data as recommended in (Tukey, 1977). For each set of data, extreme outliers, i.e., the values in the data set which were smaller than $Q1 - 3 \times IQR$ or greater than $Q3 + 3 \times IQR$ (where $Q1$ is the first quartile and $Q3$ the third quartile), were removed. Considering the gaming session data, there was one extreme outlier in both conditions for right CS data, two in the right ZM data set for the VG condition, and three in the right ZM data set for the NVG condition. Considering the watching of the first video, there was one extreme outlier in the right CS data set for the NVG condition, and two in the right ZM data set for the NVG condition. Considering the watching of the second video, there were two extreme outliers in the right CS data set for the VG condition, and two for both conditions in the right ZM data set. The right ZM data for one participant in the experiment had to be discarded due to a poor application of the sensor to the skin.

For each physiological variable, we used the Shapiro and Wilk (1965) normality test to check if data followed a Gaussian distribution. When data was normally distributed, or a mathematical transformation (logarithm or square root) could be applied to make the distribution more symmetric (Cohen, 2001), t-tests were performed. For those physiological data which could not be normalized, and for AG scores, we performed non-parametric analysis (Mann and Whitney's *U* test). To estimate the effect size for significant differences reported by t-tests, we employed Cohen's *d* (Cohen, 1988).

Finally, we carried out a correlation analysis between AQ and VGQ scores, as it has been proposed in the literature (Anderson and Dill, 2000). In the following sections, we describe the results in detail.

4.5.1 Facial EMG

Mean activities of left CS and right CS recorded during the gaming session were normally distributed. To obtain a normal distribution for left ZM and right ZM data, we applied a square root and a logarithmic transformation respectively. Figure 4.7 shows the untransformed mean values of facial EMG data recorded during the gaming session.

The difference between the mean activity of left ZM during the gaming session for the VG condition ($M = 1.78$, $SD = 2.95$) and the NVG condition ($M = 0.14$, $SD = 1.61$) was statistically significant ($t(40) = 2.23$, $p < 0.05$), and the effect size was medium to high ($d = 0.69$). The difference between the mean activity of right ZM during the gaming session for the VG condition ($M = 2.00$, $SD = 2.87$) and the NVG condition ($M = 0.36$, $SD = 1.39$) was statistically significant ($t(35) = 2.246$, $p < 0.05$), and the effect size was medium to high ($d = 0.73$). For the mean activity of left CS and right CS, the differences between VG and NVG were smaller (left CS) or negligible (right CS) and could not reach statistical significance.

Facial EMG data recorded while participants watched the violent videos showed only very small and not statistically significant differences between the VG and the NVG groups (Figure 4.8).

According to Andreassi (2007), the generally lower muscle activity in right CS and right ZM indicates that the motor cortex in the left brain hemisphere (which controls those muscles) exerted less control in the emotional experience, consistently with the fact that the right brain hemisphere (which controls left CS and left ZM) is more involved in spontaneous emotional reactions.

4.5.2 HR, BVPA, and SCL

Mean HR, mean BVPA and mean SCL recorded during the gaming session with VG and NVG as well as while participants were watching the two violent videos (Figure 4.9) were almost identical and produced

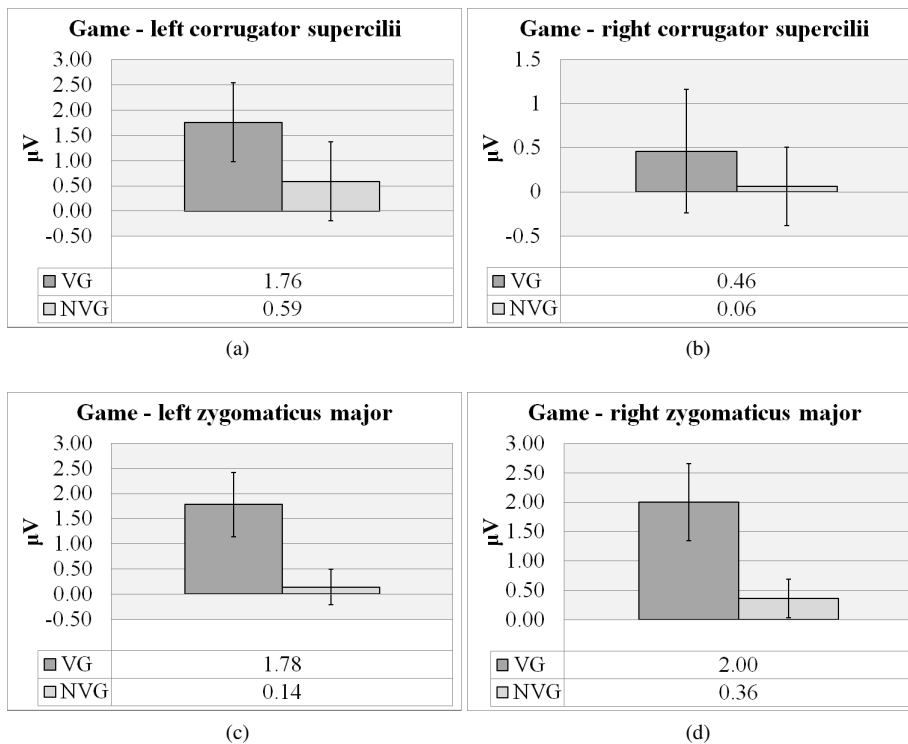


Figure 4.7: Mean CS and mean ZM activity recorded during the gaming session, after baseline subtraction. Error bars indicate standard error of the mean.

no statistically significant differences.

4.5.3 AG scores

The analysis of P score delta and A score delta, as well as video watching AG scores, found almost identical or only slightly different values between the VG and NVG groups, and none of the differences reached statistical significance.

4.5.4 Correlation analysis

Spearman's test detected a positive correlation between VGQ and AQ scores ($\rho(42) = 0.403$, $p < 0.01$), also illustrated in Figure 4.10.

4.6 Discussion of the Results

4.6.1 Players' affect during game play

During the gaming session, activity of left ZM as well as right ZM was significantly higher in VG than NVG, while no statistically significant differences were detected in left and right CS activity. This indicates that the affect elicited by the violent Whac-a-Mole game in which participants killed ants was more positively valenced than the non-violent one which replaced ants with abstract geometric shapes. This result seems to suggest that the violent element in the Whac-a-Mole game is important for the game UX, and cannot be replaced by a non-violent element in a straightforward way. A social aspect that could be considered to explain this result is that, as previously mentioned, violent behaviors against insects (and also moles, the animals used in the classic Whac-A-Mole game) tend to be socially tolerated (or in some cases even

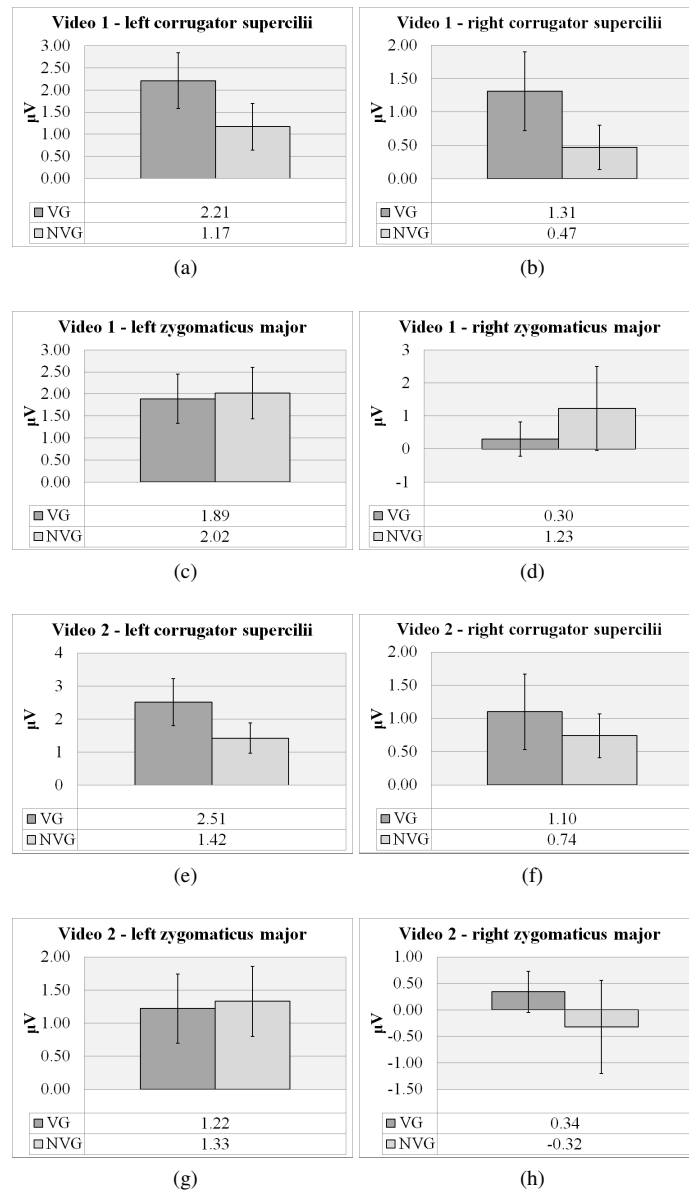


Figure 4.8: Mean CS and mean ZM activity recorded while watching the two videos, after baseline subtraction. Error bars indicate standard error of the mean.

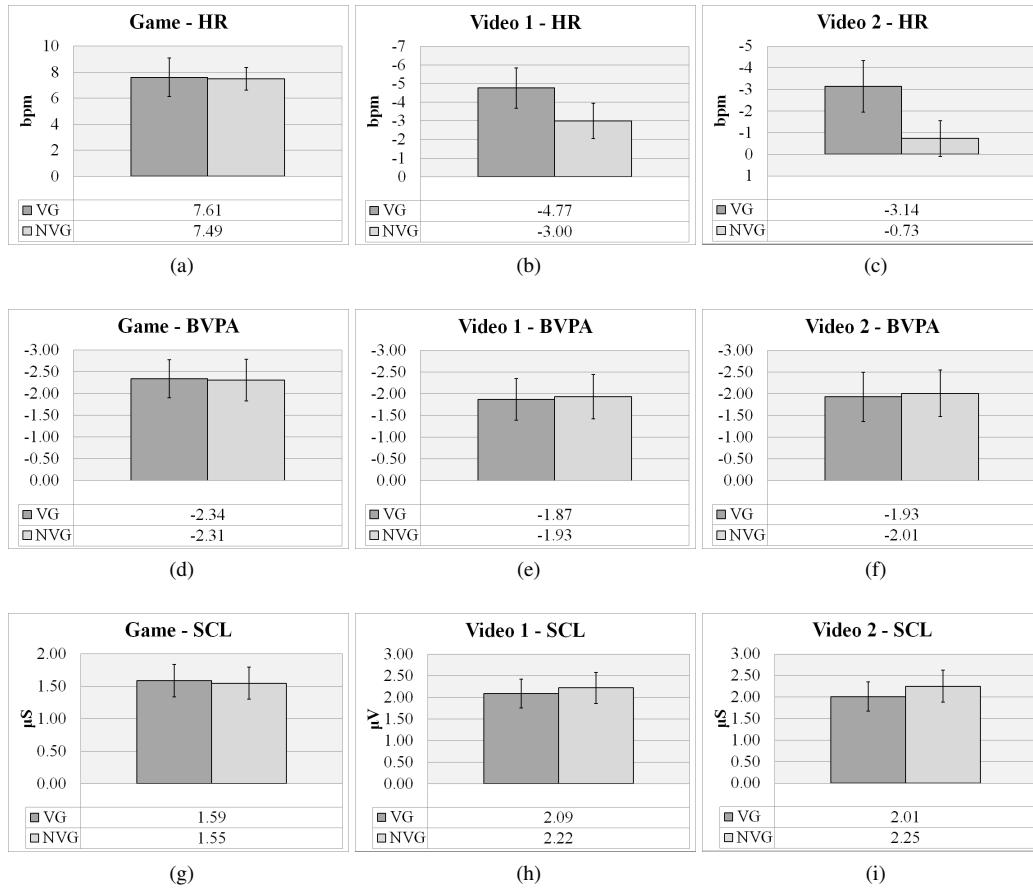


Figure 4.9: Mean HR, BVPA, and SCL values after baseline subtraction, recorded during the gaming session as well as while participants were watching the two violent videos. Error bars indicate standard error of the mean.

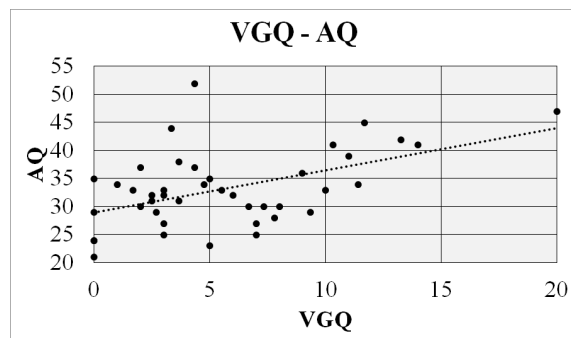


Figure 4.10: Scatterplot of VGQ and AQ scores. The regression line is also outlined.

encouraged), making them less emotionally loaded (or even pleasant). Whac-a-Mole games could thus be successful because they allow players to engage in violent acts against species that do not elicit the same physiological responses (and the corresponding affect) that have been instead highlighted in killing or wounding virtual human beings (Ravaja et al., 2008): hitting and killing species such as insects and moles is unlikely to (consciously or unconsciously) evoke taboo or moral stigma. From this point of view, it would be interesting to repeat the experiment with animal species to which people feel emotionally closer such as dogs, cats, and domestic birds.

Both games elicited instead a similar cardiovascular and EDA activity, with no statistically significant differences between the two groups. Killing ants, therefore, seems to have little effect on the arousal level elicited by the games, a different outcome compared to traditional research on the desensitizing effects of violent video games (e.g., Carnagey et al., 2007). Two factors might play a role in explaining this result: (i) the above mentioned general attitudes towards some animal species might have mitigated the arousal effects of killing insects; (ii) while the literature on violent video games has so far focused on desktop PC and console games, our participants played a game on a mobile phone, which might not be able to provide rich sensory stimulation at the intensity of a desktop system, making it possibly more difficult to emotionally arouse players. However, previous research (Arriaga et al., 2006, 2008) explored if the effects of playing violent video games (in terms of arousal and aggressive thoughts and behavior) could be mediated by the type of employed devices, considering a typical computer screen vs. a virtual reality head-mounted display (supposed to provide a stronger sense of immersion in the experience), and finding no significant differences. Since people increasingly play video games on mobile devices, it would be interesting to contrast typical desktop gaming hardware also with mobile devices to further explore if the employed devices might play a role in determining the effects of violent video games on players, as well the possible impact of the platform on the UX.

4.6.2 Desensitization to violence

The physiological data collected while participants watched violent video footages showed no significant differences between players of the violent and the non-violent Whac-a-Mole game. These results seem to suggest that repeated aggressive behavior against ants in VG did not result in different desensitizing effects to violence against insects or human beings with respect to NVG. These results differ from those in the literature, presented in Section 4.2.4, which highlighted desensitizing effects of playing violent video games. This difference could also be explained by the factors discussed in Section 4.6.1, i.e., social tolerance towards violence against animals such as insects and moles or lower level of immersion provided by mobile devices with respect to desktop PCs.

4.6.3 Correlation analysis

The positive correlation between previous exposure to violent video games (VGQ score) and aggressive traits (AQ score) is consistent with the literature on violent video games (Anderson and Dill, 2000).

4.6.4 Limitations of the study

Some limitations of our study should be considered when analyzing the obtained results. First of all, the statistical power of the study is reduced by the number of participants (42), which is small for a between-subjects study focused on aggressive behaviors and desensitization to violence (where hundreds of participants are not uncommon, e.g., Carnagey et al., 2007). The size of the data set might have hidden possible desensitization effects of the violent video game. Furthermore, while males and females were equally distributed between the two conditions, the larger number of males (32) with respect to females (10) makes the comparison of our study with the literature more difficult: in classical violent video game studies, such as those by Anderson and Bushman (e.g., Carnagey et al., 2007), gender is typically well balanced.

Finally, the violent condition of our study concerned only violence on insects. Adding a third condition to the experiment in which the game would require to aggress virtual humans might help in comparing the effects of performing aggressive actions against human and non-human animals, allowing us to better contrast our findings with the existing literature.

5

Exploring Eye–Blink Startle Response as a Physiological Measure for Affective Computing

In Chapters 3 and 4, we focused on evaluating users' affect state by exploiting the physiological signals that are most widely employed in the affective computing literature. In this chapter, we focus instead on EBSR (see Section 1.3.2), a physiological signal commonly employed in the psychophysiology literature which potential has not been yet fully exploited in affective computing applications (Chittaro and Sioni, 2013) for the detection of users' emotional states. As reported in Section 1.3.2, procedures to elicit EBSRs are often based on acoustic stimuli, in particular short bursts of intense white noise. Unfortunately, following this approach in affective computing applications would not be natural, because the artificial white noise bursts would be intrusive sounds unrelated to the user experience. They would thus distract the user from the meaningful events in the application and user's (conscious or unconscious) attempts to relate those events to the artificial stimuli would be unsuccessful and frustrating, making the stimuli detrimental to the UX. Our research aims at exploring if sounds that have a meaningful relation with the events in the application could be used as an alternative to white noise bursts. The study we present in this chapter compares physiological responses of users of a VE in two conditions (measurement of EBSR to white noise or to an alternative sound that is related to the experience), showing that the two types of acoustic stimuli produce a similar intensity of startle response. This result suggests that EBSR could be used in natural ways to extend the set of physiological measures currently employed in affective computing studies and exploited by ACSs proposed in the literature (see Chapter 2).

5.1 Introduction

As reported in Chapters 1 and 2, in affective computing detection of stress and negative emotions is typically based on measurements of EDA, facial EMG, HR, and HRV; less frequently, skin temperature, respiration and pupillometry have been considered (see Section 2.2). However, the psychophysiology literature offers additional measures, such as EBSR (Section 1.3.2), whose potential has not yet been exploited in affective computing applications.

In most studies, startle responses are generally elicited by acoustic stimuli, in particular short bursts of intense white noise at 100–110 dB. Unfortunately, following this approach in affective computing applications would not be natural, because the artificial white noise bursts would be intrusive sounds unrelated to the user experience. They would thus distract the user from the meaningful events in the application and users' (conscious or unconscious) attempts to interpret them or to associate the bursts to a particular object or event would be unsuccessful and frustrating. As a result, those stimuli would be detrimental to the user experience.

The goal of the research presented in this chapter is to assess if more realistic and natural acoustic stimuli (more specifically, sounds that can be easily recognized because they naturally presents themselves in real–world events, and can be easily interpreted in the context of the application because they have

a meaningful relation with the visual cues presented by the application itself) elicit comparable EBSRs compared to traditional white noise stimuli. This would open up the possibility to extend in a natural way the set of physiological measures employed in affective computing with EBSR.

To reach this goal, we compared the intensity of participants' EBSRs to two different acoustic stimuli, an intense burst of white noise typically used in the psychophysiology literature, and a sound with similar characteristics (in terms of duration and intensity) that is produced in the real–world by explosions. These two stimuli were presented to participants while they used a VE built with a game engine. The VE reproduces a school building during a fire emergency, a virtual experience to which explosion sounds may naturally belong.

The chapter is organized as follows. In Section 5.2 we describe how EBSRs are being used in the literature for the evaluation of VEs. The two acoustic startle stimuli as well as the VE employed in the present study are discussed in Sections 5.3 and 5.4 Sections 5.5 and 5.6 describe in detail the experiment and its results, while Section 5.7 critically discusses the results and present the future work.

5.2 Eye–Blink Startle Response and Virtual Environments

Recent psychophysiological studies resorted to VEs as a source of visual stimuli, and provided further support for the validity of EBSR as a measure of negatively–valenced emotions (see Section 1.3.2). For example, researchers have used dark immersive VEs to elicit fear and were able to measure the effect with EBSR (Grillon et al., 1997; Grillon and Ameli, 1998). In (Mühlberger et al., 2008), authors evaluated the effects of darkness while users were actively driving or sitting in the passenger seat of a virtual car that was travelling through a virtual tunnel characterized by the presence or the absence of darkness. EBSRs were able to discern the anxiety level generated by the two different VE when the user was passively immersed in the virtual driving task. Mühlberger et al. (2007) employed a virtual tunnel VE to assess the effect of tunnel phobia on startle response, observing again that EBSR is a good measure to assess fear– and anxiety–related emotional states. Immersive and realistic VEs were employed by Cornwell et al. (2006), who correlated the startle potentiation with the anxiety elicited by a socially anxious virtual experience (standing center–stage in front of an audience to anticipate giving a speech) and a less anxious one (an empty room).

EBSR was used by some researchers also as a measure of immersion in VEs (Nichols et al., 2000; Parsons et al., 2009), and a possible source of benchmarking metrics for the efficacy of VEs in therapeutic and training applications (Iyer et al., 2009). The use of devices like *head–mounted displays* (HMDs) to increase immersion in VEs seems to have a significant effect on the potentiation of EBSRs (Parsons et al., 2012).

5.3 The Selected Startle Stimuli

The white noise stimulus we use in the present study replicates the stimulus employed in (Mühlberger et al., 2008), which follows the guidelines defined by Blumenthal et al. (2005). It is a 40 ms burst of 103 dB (A) intensity with instantaneous rise time, no decay and no release time, and a flat power density over the entire audible frequency spectrum (20–20000 Hz).

The chosen natural sound is inevitably characterized by some differences with respect to the white noise stimulus. It is the sound of an explosion that reaches the same intensity of the white noise stimulus but is longer (600 ms). It is characterized by a brief rise time (about 5 ms), no decay time, a sustain time of about 200 ms, and a release time of about 400 ms. The bandwidth is in the audible range, with almost all power spectral density in the 20–16000 Hz range. The maximum intensity is in the 20–500 Hz band (with a peak in the first octave); intensity decreases linearly by 1 dB about every 1000 Hz in the 500–6000 Hz range, and by 1 dB about every 500 Hz in the 6000–16000 Hz range.

To choose the natural startle stimulus, we first collected ten sounds of explosions and/or objects crashing. This initial choice was motivated by the fact that the virtual experience reproduced by the VE was a fire emergency, a situation in which the occurrence of that kind of sounds is meaningful and appropriate. Among them, we selected those with shorter rise time and shorter total duration. Among the three selected sounds, we chose the one that sounded easiest to recognize (an explosion sound).

Startle stimuli (as well as any other sound in the VE) were presented to the user through a setup composed by two loudspeakers, placed on the sides of the computer monitor and connected to a 120 W RMS amplifier. We preferred this setup to headphones to avoid interferences with EMG physiological sensors as suggested by Blumenthal et al. (2005).

To guarantee that the intensity of the explosion sound and the white noise were the same, we measured then during the audio system set-up with a Brüel & Kjær sound level meter, placed at 1.1 m height from the floor and 1.5 m distance from the screen. The chair on which participants sat was placed in such a way that their head was located in the same position where the intensity of sound had been measured, and they could not move the chair during the task. We also measured the background sound intensity of the VE, which was designed to remain inside the 65–75 dB (A) range during the entire experience. This is a common intensity range for background sound in startle response studies (Blumenthal et al., 2005).

5.4 The Considered Virtual Environment

In our study, we employed a VE that reproduces a multi-floor school building in which each floor is made by classrooms and corridors. In the virtual experience, the user starts at the top floor, inside an empty classroom. To move inside the VE, (s)he employs a Nintendo Nunchuck controller, equipped with a joystick and two buttons. By moving the joystick forward or backwards, the user walks respectively forward and backward in the VE; by moving the joystick to the left or to the right, the user rotates his body respectively counterclockwise or clockwise in the VE. The larger of the two buttons is used to open the doors. The user is required to evacuate the building by following a series of green emergency exit signs placed over the doors and the walls to lead her to the ground floor. Only the doors belonging to the exit path could be opened, all other doors in the building were locked.

The VE reproduces a fire emergency: a constant amount of smoke fills the different areas of the environment as shown in Figure 5.1, and there are occasional corpses of victims (which the user cannot interact with) lying on the floor along the path.

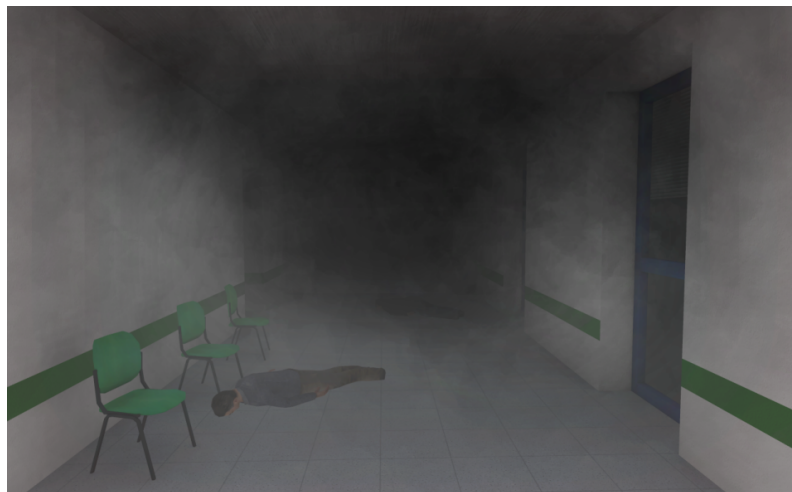


Figure 5.1: A screenshot of the VE employed in the study showing a corridor full of smoke and corpses lying on the floor.

The audio background is made of screaming people, ambulance sirens, and fire alarm sounds; the user can also hear her “virtual” breath and heartbeat. At the beginning of the virtual experience, the user is healthy, and is able to virtually run inside the VE. However, as the experience progresses, visual and auditory cues simulate her degrading health conditions caused by smoke inhalation. More specifically, she can hear herself breathing with more and more difficulty, the breathing sound becomes more and more intense; the frequency and the intensity of the played heartbeat increases; and a red aura (Figure 5.2) flashes in synch with heartbeat, progressively increasing in size to reduce the user’s field of view. Walking speed slows

down as time passes, and at the end of the virtual experience (3 minutes after the start) the user is almost unable to move.

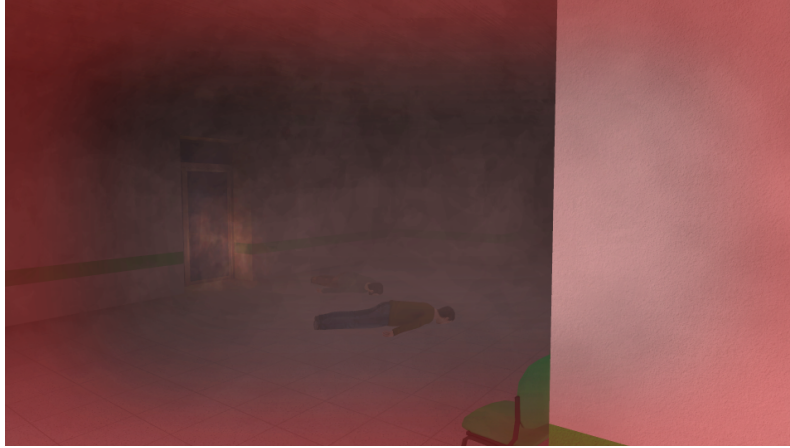


Figure 5.2: A screenshot of the VE showing the red aura.

5.5 Experimental Evaluation

The experiment followed a between–subjects design, with *startle stimulus* (*white noise* or *explosion sound*) as the independent variable and measured eye–blink response magnitude.

5.5.1 Participants

The evaluation involved a sample of 36 participants (21 M, 15 F) recruited among students. Age ranged between 16 and 35 ($M = 23.89$; $SD = 3.34$). Only one participant was a minor, and parents' approval was obtained in her case. A demographic questionnaire was employed to gather information about participants' age, gender, and video game usage. Video game usage in the sample was as follows: 4 participants never play video games; 7 participants play less than once a month; 2 participants about once a month; 8 participants more than once a month; 8 participants more than once a week; 6 participants about 1–3 hours a day; one participant more than 3 hours a day. Usage of 3D video games with realistic graphics was as follows: 11 participants never play them; 8 participants play less than once a month; 2 participants about once a month; 5 participants more than once a month; 7 participants more than once a week; one participant every day for less than one hour; 2 participants 1–3 hours a day.

The results of the demographic questionnaire were used to assign participants to the two levels of the *startle stimulus* variable in such a way that the averages of the four demographic indexes (age, gender, video game usage, realistic 3D video game usage) were as similar as possible for the two groups. More specifically, mean age was 24.11 ($SD = 4.14$) for the white noise group and 23.67 ($SD = 2.38$) for the explosion sound group; mean video game usage on the employed scale (0: never; 1: less than once a month; 2: about once a month; 3: more than once a month; 4: more than once a week; 5: every day, for less than 1 hour; 6: every day, for about 1–3 hours; 7: every day, for more than 3 hours) was 2.94 ($SD = 1.76$) and 3.16 ($SD = 2.31$) respectively; mean usage of 3D video games with realistic graphics was 1.89 ($SD = 1.81$) and 2.11 ($SD = 2.03$) respectively.

5.5.2 Materials

The two VEs were run on a Mac Pro (with two quad–core processors at 2.8 GHz; 4 GB RAM; NVidia 8800 GT graphic card with 512 MB of dedicated memory) on a 30" LCD monitor at WQXGA resolution (2560 × 1600 pixels).

To record participants' EBSR data, we employed an EMG sensor with two disposable pre-gelled electrodes attached at a constant inter-electrode distance of 2 cm over the orbicularis oculi muscle, beneath the left eye, following the electrode placement recommended by Fridlund and Cacioppo (1986) (Figure 5.3).



Figure 5.3: The two black dots pinpoint the location of EMG electrodes for EBSR measures.

The signal, recorded using a ProComp Infiniti encoder, was sampled at 1024 Hz. Following the guidelines by Blumenthal et al. (2005), a band-pass filter in the 28–500 Hz range was first applied, then the filtered signal was rectified with a moving window and finally smoothed with a variable-weight FIR filter with a low-pass cutoff frequency of 40 Hz.

5.5.3 Procedure

Participants were verbally instructed about the task, and informed about the presence of intense but very short noises (the startle stimuli) during the experience. They were also clearly informed that all the experimental data was going to be collected and analyzed anonymously for research purposes. The white noise and the explosion sound were presented to them two times each. Then, participants were asked if they agreed to participate to the experiment. All of them consented to participate and none of them found the stimuli intensity intolerable. After they filled the demographic questionnaire, participants were seated and the skin under their left eye was cleaned using a pad of cotton wool and alcohol, then the EMG sensor was applied.

Then, they were asked to relax for a minute while watching a video with relaxing images and music, so that physiological parameters could reach a rest state. Finally, they carried out the task, which lasted 3 minutes.

During the task, six startle stimuli were delivered to participants. The first stimulus was administered $29 \text{ s} \pm 5 \text{ s}$ after the start of the experience (the value in the $\pm 5 \text{ s}$ range was randomly chosen). Similarly, the distance between each pair of subsequent stimuli was $29 \text{ s} \pm 5 \text{ s}$, with the random numbers chosen in such a way that the mean of the five temporal distances between two consecutive stimuli was 29 s. The purpose of this pseudo-randomization of the startle stimuli timing is to prevent learning effects.

Finally, participants were debriefed about the experiment and thanked for their participation.

5.6 Results

5.6.1 Data reduction and analysis

To gather EBSR data, we followed a procedure similar to the one described in (Mühlberger et al., 2008). We analyzed manually the filtered EMG data, and identified the EMG value of the signal peak in a 20–500 ms window after the stimulus. The width of this time window has been set to preemptively take into account any possible latency caused by EMG signal filtering, as recommended in (Blumenthal et al., 2005). The window can also take into account possible latency that might be introduced by the slightly longer duration

and rise time of the employed explosion sound startle stimuli. An example of EBSR (filtered as described in Section 5.5.2) is provided in Figure 5.4.



Figure 5.4: Filtered EMG signal showing an EBSR to a stimulus presented in correspondence to the dot-dashed vertical line.

The baseline value for each startle response was calculated as the mean EMG value in the 0–20 ms time interval after the startle stimulus presentation. The EMG signal was visually examined to detect possible situations in which artifacts might have negatively affected the assessment of a startle response. This may happen, for example, if the user closes her eyes during the startle stimulus presentation and, as a result, the baseline for determining the startle magnitude becomes unreliable. In these (relatively rare) cases, the affected startle response has to be discarded. To obtain the magnitude for a startle response, the relative baseline value was subtracted from the EMG value of the peak.

One participant in the white noise group had to be excluded because her EMG data was corrupted by a malfunctioning or improperly applied disposable electrode.

For each participant, we calculated the mean EMG value of the six startle responses. We excluded outliers by using the method recommended in (Blumenthal et al., 2005): participants with a mean EMG value larger than 3SD from the mean EMG value of the group they belong to were removed from the data set. As a result, we excluded two participants from the analysis, one for each group.

5.6.2 Results

Before analyzing the data set with a two-tailed unpaired *t*-test, we assessed data normality (a prerequisite for the *t*-test) using the Shapiro–Wilk test (Shapiro and Wilk, 1965). The results showed that the data did not follow a Gaussian distribution; therefore, we applied a log-transformation, as suggested in (Cohen, 2001), to normalize the data.

The *t*-test performed on the transformed data revealed no significant differences ($t(31) = 1.15$, $p = 0.26$) between the white noise group ($M = 9.35$, $SD = 6.70$) and the explosion sound group ($M = 7.84$, $SD = 7.84$). The untransformed mean values are shown in Figure 5.5.

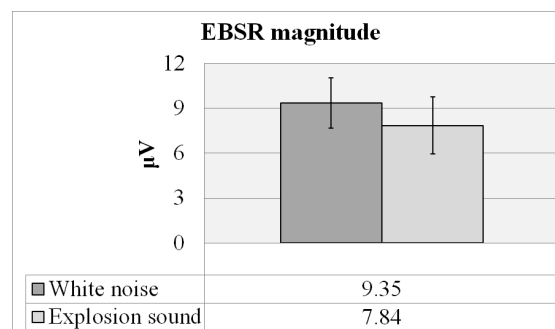


Figure 5.5: Mean magnitude of EBSRs to the white noise and the explosion sound. Error bars indicate standard error of the mean.

5.7 Discussion of the Results

In the present study, we compared EBSRs elicited by a traditional white noise startle stimulus and a more natural explosion sound of equal intensity. Participants were immersed in a VE reproducing a school building during a fire emergency, a virtual experience to which explosion sounds may naturally belong.

The results of the study show that: (i) the explosion sound is able to elicit an EBSR, and (ii) the magnitude of the response is close to the one elicited by white noise, and the small difference is not statistically significant. This indicates that the rise time of the explosion sound (one of the parameters of acoustic startle stimuli that seems to play the most important role in startle potentiation (Blumenthal et al., 2005)) was short enough to elicit an intense startle response. Indeed, one of the factors that led us to choose that particular explosion sound was its short rise time (about 5 ms). Similarly, the longer duration of the natural noise stimulus did not result in a significant negative impact on the magnitude of the startle response. While the white noise is shorter and maintains its intensity for the entire duration of the stimuli, the explosion sound is longer but the decay described in Section 5.3 actually reduced the time during which it is intense, attenuating the possible negative effects of the longer duration. In any case, we could not find in the literature studies comparing the effects of different stimulus durations. A 40 or 50 ms duration is typically employed and sufficient for startle elicitation. Longer white noise stimuli are probably not employed because, while not unsafe, they may be too uncomfortable for participants (Blumenthal et al., 2005).

6

Physiology–Based Evaluation of a Mobile Breathing Training Application

Training is one of the most important areas of application of affective computing, as reported in Section 2.3.1. With this chapter, we start to explore this topic, by applying the knowledge gained with the studies described in the previous chapters for the evaluation of a breathing training mobile app. Deep and slow breathing exercises have positive effects on blood pressure and hypertension and can be an effective adjunct in the treatment of stress, anxiety, post-traumatic stress disorder, chronic pain and depression. Mobile breathing training apps, therefore, may play a significant role in relaxation training activities. Breathing techniques are traditionally learned in courses with trainers and/or with materials such as audio CDs for home practice. Recently, mobile apps have been proposed as novel breathing training tools, but to the best of our knowledge no research has evaluated their effectiveness so far. In this chapter, we study three different designs for breathing training apps. The first employs audio instructions as in traditional training based on audio CDs, while the other two include visualizations of the breathing process, representative of those employed in current breathing training apps. We carry out a thorough analysis, focusing on users' physiological parameters as well as subjective perception. In particular, we focus on "traditional" physiological signals (EDA, HR, respiration) as in the study described in Chapter 4, instead of novel signals like the EBSR (Chapter 5), which may be unsuitable for the present application. One visualization produced better results both objectively (measured deepness of breath) and subjectively (users' preferences and perceived effectiveness) than the more traditional audio-only design. This indicates that interactive visualizations are able to improve the effectiveness of breathing training apps. We discuss which features could have allowed one visualization (but not the other) to outperform traditional audio-only instructions.

6.1 Introduction

Deep and slow breathing, as well as more complex breathing exercises such as Yogic breathing, have been found to be an effective adjunct in the treatment of stress, anxiety, *post-traumatic stress disorder* (PTSD), chronic pain and depression (Brown and Gerbarg, 2005; Busch et al., 2012; Han et al., 1996), and in the achievement of relaxation (Grossman et al., 2001). These benefits can be explained by the fact that deep and slow breathing exercises contrast the effects of fast and shallow chest breathing, which is a common automatic habit developed by patients with anxiety disorders (Hazlett-Stevens, 2008) and has been associated to physiological effects like hyperventilation, which could lead to physical sensations resembling anxiety (Hazlett-Stevens, 2008) and respiratory symptoms typical of panic attacks (Conrad et al., 2007). Furthermore, several medical studies have reported that breathing exercises can also have positive effects on the circulatory system, by helping to lower blood pressure (Grossman et al., 2001; Radaelli et al., 2004; Joseph et al., 2005). Integrating breathing exercises with more general relaxation training can also ameliorate respiratory symptoms in patients with asthma (Holloway and West, 2007).

Breathing exercises are traditionally learned in courses with trainers, e.g., courses for pregnant women in preparation to labor or as part of programs for dealing with medical conditions such as hypertension

or stress-related disorders. Audio CDs are often used to support breathing practice at home. Recently, some authors (Schein et al., 2001; Elliott et al., 2004; Mitchell et al., 2010; Liu et al., 2011) have focused on specialized hardware to guide the trainee during breathing exercises. In general, these systems exploit wearable devices like elastic girth sensors to record users' breathing activity (e.g., Mitchell et al., 2010; Liu et al., 2011), sometimes combining them with other devices, e.g., a portable music player (e.g., Schein et al., 2001; Elliott et al., 2004). The goal of these systems is to provide real-time adaptation of the training to users' physiology. Unfortunately, specialized hardware may be costly and not always available when required, e.g., when stress strikes. Furthermore, the transition to applying the skills learned through exercises in everyday life remains a challenge (Morris and Guilak, 2009).

Smartphones can be a novel opportunity to improve breathing training. They follow their users anywhere, so a breathing training app could be always available to support users at any moment. Moreover, the cost of a smartphone app is typically low, and is not infrequent to find free apps. In general, smartphones are increasingly seen as a versatile m-health instrument for treatment and training, in medicine as well as psychology (Miller, 2012), and some authors predict that the mobile phone will emerge as the preferred personal coach for the 21st century (Morris and Guilak, 2009). A few mobile apps for breathing training have been proposed in the literature, e.g., in the Mobile Heart Health project (Morris and Guilak, 2009). Moreover, app stores such as Apple's App Store and Google Play Store are making available a growing number of mobile apps for breathing training, developed both by small enterprises (e.g., Universal Breathing: Pranayama (Saagara, 2011)) and organizations such as the US National Center for Telehealth and Technology (T2), a part of the US Military Health System (MHS), which has recently launched Tactical Breather (National Center for Telehealth and Technology, 2011b), a mobile app for repetitive breathing training. The goal of Tactical Breather is to help soldiers in gaining a better control of heart rate, emotions, concentration, and other physiological and psychological responses during stressful situations (National Center for Telehealth and Technology, 2011b).

Unfortunately, to the best of our knowledge, no research study has yet focused on evaluating the effectiveness of mobile breathing training apps. In this chapter, we study three different designs for such apps. The first employs audio instructions as in traditional training based on audio CDs. The other two include also visualizations of the breathing process, which are representative of approaches followed in current breathing training apps. We carry out a thorough analysis of participants' physiological parameters as well as their subjective perception and preferences. In particular, we focus on "traditional" physiological signals (EDA, HR, respiration) as in the study described in Chapter 4, instead of novel signals like the EBSR (Chapter 5), which may be unsuitable for the present application.

This chapter is organized as follows. Section 6.2 briefly reviews the various apps for breathing training, while Section 6.3 provides additional motivations for our research. In Section 6.4, we describe the three designs considered in the present study. Then, Sections 6.5 and 6.6 illustrate in detail the experiment and its results, while Section 6.7 discuss the results.

6.2 Related Work

Visualization plays an important role in many kinds of medical applications (Chittaro, 2001) and is increasingly employed in mobile devices to make the information provided by their applications easier to understand (Chittaro, 2006). While some breathing training apps rely on audio-only instructions (as in the traditional approach based on audio CDs), more innovative apps are attempting to enrich audio instructions with interactive visualizations. Some of these apps share a common approach (called *circle-based visualization* in the following): they display a circle (or a sphere) that expands and contracts, which might suggest the expansion and contraction of human lungs during the breathing activity. For example, in (Morris and Guilak, 2009; Morris et al., 2010) a simple blue circle is animated to encourage deliberate and slower breathing (Morris et al., 2010). To better highlight the different phases of breathing, Tactical Breather (National Center for Telehealth and Technology, 2011b) also changes the color of the circle. During the inhalation phase, the circle is green and grows in size; in the hold phase (during which users must hold their breath), the circle turns yellow and its size remains constant; in the exhalation phase, the circle turns red and shrinks until it reaches its minimum size, which is marked by a light black circle in the center of the screen. During each phase, voice instructions first pronounce the name of the phase ("Inhale", "Hold" or

“Exhale”), then count from 2 to 4 to give users an indication of the progress and duration of the phase. The phase and the pronounced number are also displayed as text on the screen. ColorBREATH (3CUBEs, 2012) substitutes the solid color circle with an iridescent bubble. ECNA–Breath (ECNA LAB, 2011) employs a green sphere that expands and contracts, but it allows users to adjust the maximum size of the sphere before starting the exercise to better fit it with their actual breathing depth. Breathing Zone (Breathing Zone, 2011) employs a multicolor geometric shape similar to a lotus flower instead of a circle. Two different sounds, that can be chosen among various tones (e.g., two guitar chords) announce the beginning of the inhalation and exhalation phases.

Some apps follow a different approach (called *wave-based* visualization in the following), employing a wave-like line to show the optimal respiratory cycle. For example, Vital–EQ Respiroguide (Landelijk Centrum Stressmanagement, 2009) shows the breathing pattern as a sine wave that advances over time during the exercise. A yellow circle, locked on the curve and at the center of the screen, moves upwards and downwards following the wave to guide the user through the inhalation and exhalation phases. A wave-like line is also employed in Paced Breathing (IQPuzz, 2013): in this app, the wave represents a single respiratory cycle and does not move. During the exercise, a white circle follows the line to indicate the current position in the respiratory cycle. A tone with increasing or decreasing pitch is employed as an audio hint to indicate inhalation and exhalation respectively. The shape of the wave in terms of the length of breathing phases, but not their amplitude, is customizable by the user.

Two apps, Universal Breathing: Pranayama (Saagara, 2011) and TotalAwake (TotalAwake, 2012), have tried pie charts instead of waves. Pie sectors represent the different breathing phases and during a respiratory cycle, the chart is gradually filled with color to guide users through the different phases. Breathe2Relax (National Center for Telehealth and Technology, 2011a) employs a cylinder shape which fills up and empties out to indicate when to inhale and exhale. De Stress (Designit, 2009) employs instead animated arrows, which move towards or away from the screen to visually represent the inhalation and exhalation phases. The arrows move with decreasing speed during the exercise to help users gradually slow down their breathing frequency.

Finally, three apps resort to visualizations of the human body. Universal Breathing: Pranayama and Paced Breathing display a realistic 3D model of a human in the so-called “Burmese” posture, animated to represent the expansions and contractions of chest and abdomen during the exercise. A small blue arrow indicates the air coming in and out through the nose of the human. MyCalmBeat (MyBrainSolutions, 2012) introduces a silhouette of a human chest that encloses two 3D lungs. An animation shows when to inhale or exhale: a tone indicates the start of the inhalation phase; the 3D lungs expand and get gradually filled with oxygen, represented by a blue shade. When the lungs are full, a tone indicates the start of the exhalation phase, and the animation is reproduced in reverse order.

6.3 Motivations and Goals

The lack of studies of the effects of mobile breathing training apps on users does not allow one to make substantiated claims about the effectiveness of the above described visualizations. On one side, visualizations may offer easier to understand instructions and help trainees in reaching the optimal breathing pattern. On the other side, they could possibly distract users’ attention from respiratory proprioception which is important in controlling breath, and be detrimental to reaching slow and deep breathing. Their effects on breathing must thus be studied and contrasted with the traditional audio-only approach. For this reason, our study will consider the two currently most used types of visualization (summarized in the previous section) as well as the more traditional audio-only approach, to test if the inclusion of a visualization can improve the latter.

Moreover, we should consider that mobile breathing visualizations differ in the amount of information conveyed to users. The most limited ones only suggest the current phase of the respiratory cycle, as it is the case of those circle-based visualizations which convey information only through circle contraction and expansion, without giving an explicit indication of how long the user is supposed to keep inhaling or exhaling. Some apps, like Tactical Breather, try to overcome this limitation by providing additional visual and acoustic hints such as counting the duration of each phase, or marking the minimum size of the circle, i.e., the moment when users will have to stop exhaling. However, these suggestions may not be detailed

and clear enough for novice users who have just begun practicing deep and slow breathing. From this point of view, some wave-based visualization such as Vital-EQ Respiroguide might be more clear, because they provide users with temporal information not only about the current breathing phase but also about the previous and the following phase, possibly allowing users to gain greater awareness of where they are in the planned respiratory cycle and thus foresee with greater precision when to switch between phases. For this reason, we hypothesize that wave-like visualizations could allow users to modulate more easily their breathing pattern and better match it with the suggested pattern. Our study is thus also meant to provide guidance in the choice between circle-based and wave-based visualizations.

6.4 The Considered Designs

For our study, we developed three versions of a mobile app for the Android platform that differ only in the way instructions are presented. The app guides users in practicing deep and slow circular breathing: a breathing cycle is composed of two phases (inhalation and exhalation) of the same length and the breathing frequency is 6 cycles per minute (0.1 Hz). As in many breathing training apps, the first version of the app (called *Voice-only* in the following) provides instructions through audio only. In particular, we employ the same voice instructions of Tactical Breather, i.e., “inhale”, “2”, “3”, and “4” for the inhalation phase, and “exhale”, “2”, “3”, and “4” for the exhalation phase. The time between any two consecutive instructions is set at 1.25 s.

The second version of the app augments *Voice-only* with a circle-based visualization. This version of the app (called *Sphere* in the following) was implemented to replicate the visualization used for inhalation and exhalation in Tactical Breather: it employs a green inflating sphere (see Figure 6.1) to instruct users to inhale, and a red deflating sphere to instruct them to exhale. As in Tactical Breather, a light black circle is shown at the center of the screen to indicate the minimum size of the sphere and the audio instructions are also visually presented: numbers are shown in the center of the sphere (“1” is simultaneous to the voice instruction “inhale” or “exhale”, while “2”, “3”, “4” appear simultaneously to their audio versions); the textual indication about the current breathing phase (“inhale” or “exhale”) is shown at the bottom of the screen. Two screenshots of the visualization are shown in Figure 6.1.

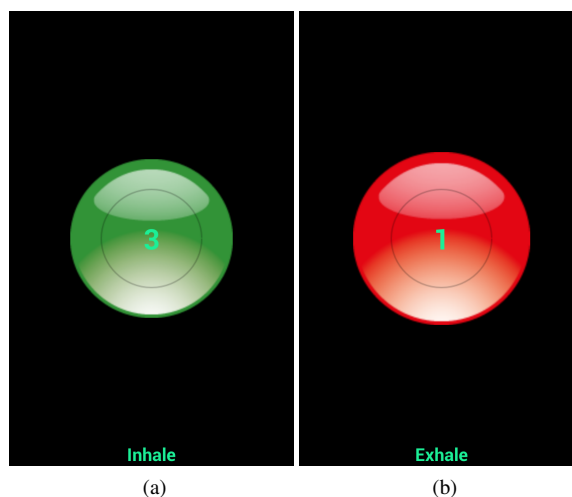


Figure 6.1: Two screenshots from the Sphere design during inhalation (a) and exhalation (b) phases, in which the sphere is green and red respectively.

The third version of the app augments *Voice-only* with a wave-based visualization. This version of the app (called *Wave* in the following) employs a moving green triangle wave (see Figure 6.2) to instruct users about when to inhale (the wave is rising) and exhale (the wave is falling), similarly to existing apps like Vital-EQ Respiroguide and Paced Breathing. A vertical translucent white strip, at the center of the screen,

is superimposed over the wave to indicate where the user should be in the respiratory cycle at the current instant of time. The previously described audio instructions are also visually presented in the top left corner of the screen. As shown by Figure 6.2, during an exhalation (respectively, inhalation) phase, users can see the falling (rising) phase of the triangular wave and also the following rising (falling) phase and the previous rising (falling) phase of the wave.

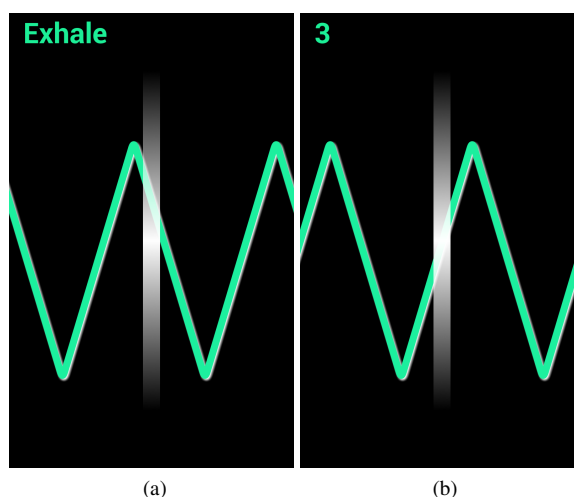


Figure 6.2: Two screenshots from the Wave design during the exhalation (a) and inhalation (b) phases.

6.5 Experimental Evaluation

Our experiment follows a within-subject design with *app version* (*Wave*, *Sphere*, and *Voice-only*) as the independent variable. We measure possible differences in participants' physiological responses during the exercises, as well as in their subjective perceived effectiveness of the three app designs.

6.5.1 Materials

The app was run on an HTC Desire equipped with a 3.7", 480 × 800 pixels display. During the evaluation, the device was placed on a tilted rigid stand placed in a preset position, to maximize visual clarity by avoiding issues like glare and reflections from light sources. Participants' physiological data was recorded and processed on a PC, whose screen could not be seen by participants. To record participants' physiological data, we employed a clinical system (Thought Technology Procomp Infiniti) equipped with 3 sensors: (i) an EDA sensor placed on the intermediate phalanges on the middle and little fingers; (ii) a PPG sensor placed over the distal phalanx of the index finger; (iii) an elastic girth sensor, placed over the participant's navel, to correctly record diaphragmatic respiratory activity.

A demographic questionnaire asked participants about their age and gender, as well as their experience with relaxation and meditation techniques, as suggested in (Feldman et al., 2010). To measure participants' perceived effectiveness of each app version, we employed a questionnaire with the following 6 items: (i) The app facilitates relaxation; (ii) The app is pleasant to use; (iii) It is easy to follow the app instructions; (iv) The app effectively teaches how to breath; (v) The app is effective in reducing stress; (vi) The app is effective in increasing attention to breath. Each item was rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The reliability of the scale was assessed with Cronbach's alpha (0.83, 0.84, and 0.84 for Wave, Sphere and Voice-only, respectively).

Participants' rating of the first and fifth items were averaged to assess the effectiveness of each app version in relaxing the participant. The reliability of the measure was assessed with Cronbach's alpha (0.81, 0.80, .and 0.83 for Wave, Sphere and Voice-only, respectively). Finally, we asked participants to

rank the three app versions, on the basis of which they would prefer to use (1 = best, 3 = worst). Ties were allowed.

6.5.2 Participants

The considered sample involved 68 participants (43 M, 25 F), recruited among graduate and undergraduate students of our university and people from other occupations. Age ranged from 17 to 60 ($M = 24.73$, $SD = 5.57$). All participants had little or no experience with breathing techniques and were thus representative of users who might be supported by a training app. More specifically, the majority of participants (55) was completely unfamiliar with exercises that involve breath control (relaxation or meditation), while 13 participants had some familiarity with such exercises but did not practice them.

6.5.3 Procedure

Participants were informed that the goal of the study was to evaluate three different versions of a mobile app that would guide them in practicing breathing exercises. All participants consented to participate in the experiment and to have physiological signals recorded. They were also clearly informed that all the experimental data were going to be collected and analyzed anonymously for research purposes. After they filled the demographic questionnaire, the girth sensor was applied accurately over their navel to measure diaphragmatic breathing. Then, they were asked to sit, and the PPG and EDA sensors were applied.

Before each condition, participants were asked to relax (the request was deliberately generic and did not mention breathing) for about a minute, to record the baseline for the physiological signals, i.e., the signal values that can be observed when participants are in a resting state. When analyzing data, baseline values have to be subtracted from the data recorded during the experimental conditions, to separate the physiological responses to experimental conditions from the intrinsic physiological differences among participants (Andreassi, 2007). During baseline recording, participants watched a video with nature sounds and images. After each baseline recording, we introduced them to the app version they were going to try next. We clearly explained that, during the exercise, they would have to breathe with the abdomen. Then, participants practiced breathing training with the given app for 3 minutes. To counterbalance learning effects, the order of the three conditions was varied.

After trying all three designs, participants filled the perceived effectiveness questionnaires, and then gave the final ranking. They were then debriefed and thanked for their participation.

6.5.4 Measures

For each participant, we collected the following measures:

- *SCL*: the mean value of the SCL measured during each breathing exercise, after the subtraction of the mean baseline value;
- *HR*: the mean value of HR measured during each breathing exercise, after the subtraction of the mean baseline value;
- *Power of the recommended frequency band*: the signal power in the 0.09–0.11 Hz band of the respiratory signal power spectrum, measured during each breathing exercise. This measure represents the intensity of the respiratory component near the breathing frequency encouraged by the app (0.1 Hz, i.e., 6 cycles per minute). Higher values for this measure suggest deeper breaths at the suggested frequency (which is the goal of the deep and slow breathing exercises);
- *Respiratory signal-to-noise ratio (SNR)*: the ratio between (i) the power of the recommended breathing frequency band and, (ii) the power in the entire spectrum excluding the band of the recommended frequency and the 0–0.05 Hz band (to remove low-frequency fluctuations). It describes the ability of participants to correctly follow the instructions received by the app: the more the participant struggles in following the instructions or ignores the instructions, the more the power of the components outside the recommended breathing frequency band, especially high-frequency components, increases.

For example, a participant who is unable to inhale further before the inhalation phase is complete might perform a brief exhalation to keep inhaling until the end of the phase;

- *Perceived effectiveness*: the mean score from the questionnaire collecting participants' subjective perception of the effectiveness of each app design;
- *Perceived relaxation effectiveness*: the mean score from the first and fifth items from the questionnaire, focusing about perceived relaxation effectiveness with each app design;
- *Subjective preference*: rank assigned by participants to each design.

6.6 Results

We performed a multivariate analysis of variance (MANOVA) of the physiological data and questionnaire results. To reduce the risk of type I errors, we adopted a conservative alpha level, setting it at 0.0083 (Bonferroni correction) for subsequent univariate tests and pairwise comparisons. A separate non-parametric analysis of variance (Friedman test) was performed on the subjective preference data. Physiological data for the respiratory activity recorded from 5 participants had to be removed due to a malfunctioning girth sensor during a day of the recording sessions. HR data from 9 participants had to be removed because of a large number of artifacts in the BVP signal due to hand movements during the recording session.

The analysis revealed a main effect of visualization (Wilks' $\lambda = 0.68$, $F(2, 106) = 3.52$, $p < 0.001$, $\eta_p^2 = 0.17$). Univariate tests with Greenhouse–Geisser correction showed significant differences in power of the recommended frequency band ($p < 0.0083$, $\eta_p^2 = 0.10$), perceived effectiveness ($p < 0.001$, $\eta_p^2 = 0.17$), and perceived relaxation effectiveness ($p < 0.0083$, $\eta_p^2 = 0.09$) (Figure 6.3a, 6.3c, 6.3d).

Differences in Respiratory SNR, SCL, and HR did not reach significance.

Pairwise comparisons (with Bonferroni correction) for power of the recommended frequency band revealed a significant difference between the Wave ($M = 10.78$, $SD = 10.46$) and the Voice-only ($M = 7.98$, $SD = 8.18$) conditions ($p < 0.0083$). Similarly, pairwise comparisons among perceived effectiveness scores revealed a significant difference between the Wave ($M = 3.85$, $SD = 0.69$) and the Voice-only ($M = 3.18$, $SD = 0.91$) conditions ($p < 0.001$), and pairwise comparisons among perceived relaxation effectiveness scores revealed a significant difference between the Wave ($M = 3.56$, $SD = 0.88$) and the Voice-only ($M = 2.96$, $SD = 1.22$) conditions ($p < 0.0083$).

Finally, Friedman test revealed a significant difference in subjective preference among the three conditions ($\chi^2 = 14.84$, $p < 0.001$) (Figure 6.4), and a pairwise comparison with Dunn's test (with Bonferroni correction) revealed a significant difference between the Wave ($M = 1.59$, $SD = 0.74$) and the Voice-only ($M = 2.24$, $SD = 0.83$) conditions ($p < 0.001$).

6.7 Discussion of the Results

The experiment showed that the Wave design produced significantly deeper breaths at the suggested frequency than Voice-only. This is probably due to the fact that Wave allows users to keep better track over time of the ideal breathing pattern they have to follow during the exercise. The visualization provided by Wave could have allowed participants to assess more easily the duration of the breathing phases, thus helping them to maximize the quantity of air to inhale or exhale in each respiratory cycle with respect to Voice-only. On the contrary, receiving only audio instructions as in Voice-only could have resulted in a more shallow breathing in participants to avoid finding themselves unable to further inhale (or exhale) before the end of the inhalation (exhalation) phase.

Although Sphere also produced deeper breathing at the suggested frequency with respect to Audio-only (Figure 6.3a), the difference did not reach significance. However, it is interesting to note that the average obtained by Sphere is only slightly smaller than Wave. The presence of the light black circle and the inflating and deflating sphere which can be associated with human lungs' expanding and contracting might have partially helped participants in calibrating their inhalation and exhalation.

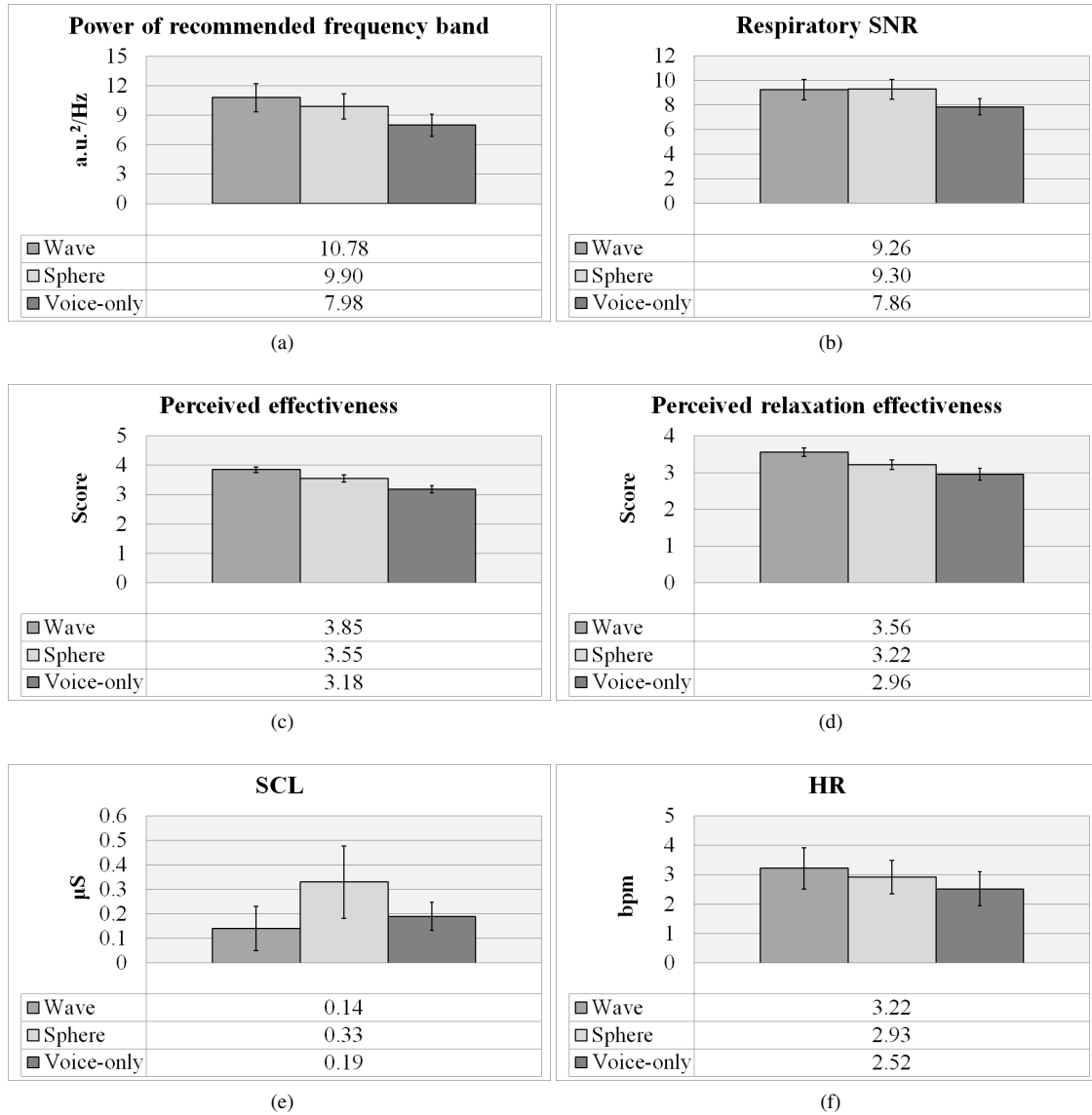


Figure 6.3: Mean values of power of the recommended frequency band (a), respiratory SNR (b), perceived effectiveness (c), perceived relaxation effectiveness (d), SCL (e), HR (f). Error bars indicate standard error of the mean. a.u. indicates the arbitrary units provided by the elastic girth sensor to measure its tension.

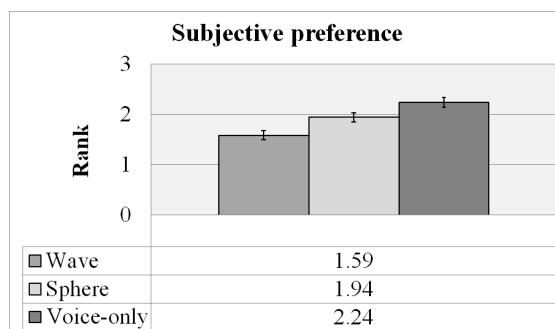


Figure 6.4: Mean values of subjective preference. A lower value indicates higher preference. Error bars indicate standard error of the mean.

The same considerations apply to results about perceived effectiveness, perceived relaxation effectiveness and participants' subjective preference, which were consistent with objective respiratory results: perceived effectiveness (Figure 6.3c) and perceived relaxation effectiveness (Figure 6.3d) were significantly higher with Wave rather than Voice-only, and participants reported a significantly higher preference for Wave over Voice-only (Figure 6.4).

The implications of the obtained results are twofold. First, including an interactive visualization in breathing training apps to help trainees in gaining greater awareness of the breathing process can have positive effects in obtaining deep and slow breathing, and can be in particular more effective than traditional audio-only instructions. Without the results provided by our evaluation, one could not exclude, as noted in the introduction, that visualizations might distract users' attention from the respiratory proprioception required to carry out breathing exercises, and thus be detrimental to the training. Second, our study also sheds light on differences between the two currently most used types of visualizations in breathing training apps. All statistically significant differences from Audio-only were obtained with the Wave visualization, which performed consistently better. Even extending consideration to the measures for which significance was not reached, Sphere and Audio-only never show better values than Wave.

7

The Proposed Affective State Detection System: ACME

After using physiological signals to evaluate the training application we proposed in Chapter 6, we build upon the literature to present an ACS called *Affective Computing MachinE* (ACME), developed to detect users' stress level from the real-time analysis of facial EMG, HR, and EDA. In this chapter, we describe in detail the ACME architecture (Section 7.2) as well as its user interface (Section 7.3). Then, in Section 7.4, we present a preliminary evaluation of its effectiveness, based on physiological data sets recorded during the experiment described in Chapter 6, as well as during a pilot study carried out expressly for this evaluation.

7.1 Introduction

ACME was inspired by Mandryk and Atkins's work (2007). It employs FL (see Section 2.1.1) to infer emotions from physiological data, employing the Mamdani inference system (Mamdani and Assilian, 1975). We have chosen an approach based on FL over other techniques, such as SVMs (Section 2.1.5), because it can easily represent continuous processes, like changes in the affective state, that are not easily broken into discrete segments, and in which the change of state from one linguistically-defined level to the next is not clear (Cox, 1992). Furthermore, as reported in the Introduction of this thesis, the ability to employ complex fuzzy rules allows FL to manage multiple fuzzy sets at once and simulate the entanglement of multiple physiological signals, unlike other techniques such as Naïve Bayes classifiers (Section 2.1.4).

As reported in Section 2.2, other ACSs proposed in the literature employ FL. In (Rani et al., 2007), authors extract four features from the frequency spectrum of IBI detected through ECG: two indices for stress and relaxation in the parasympathetic range, and two indices for stress and relaxation in the sympathetic range (see Section 1.3.3). To minimize the latency due to the ECG analysis in the frequency domain, authors employ a *Wavelet Packet Decomposition*. The two ranges of sympathetic and parasympathetic activity are defined, for each participant, by two pairs of variables that authors calculate by collecting multiple data sets on each participant. These four values are then used to define the extremes of the four input fuzzy sets. The output signal is expressed by three fuzzy sets: *least stress*, *medium stress* and *high stress*. The fuzzy inference process is implemented using four fuzzy rules, and the aggregate output is defuzzified to get an output (*degree of stress*) in the 0–1 range.

In (Katsis et al., 2008), authors employ an *Adaptive Neuro-Fuzzy Inference Systems* (ANFIS), i.e., a fuzzy system that is trained by a learning algorithm derived from neural network theory. The proposed system employs an array of wearable sensors (EMG, ECG, respiration, EDA) that wirelessly communicate with a centralized module, which performs feature extraction and perform the fuzzy inference process to classify, nearly in real time, indices of low stress, high stress, disappointment, and euphoria. The centralized module employs two fuzzy rules and four fuzzy sets, which are modeled according to the ANFIS architecture (Jang et al., 1997); Finally, a user interface module shows users the acquired signals and the emotion indices.

7.2 ACME Architecture

With ACME, we improved upon the original system in several ways:

- Raw physiological data is now read in real time from sensors;
- Baseline recording is introduced, so that data normalization and analysis can be now performed in real time;
- The algorithms for data filtering and physiological feature extraction have been updated to be able to work in real time;
- A communication module, based on TCP/IP protocol, has been introduced to let ACME interact with external applications such as virtual training systems.

In the following, the architecture of the system (Figure 7.1) is presented.

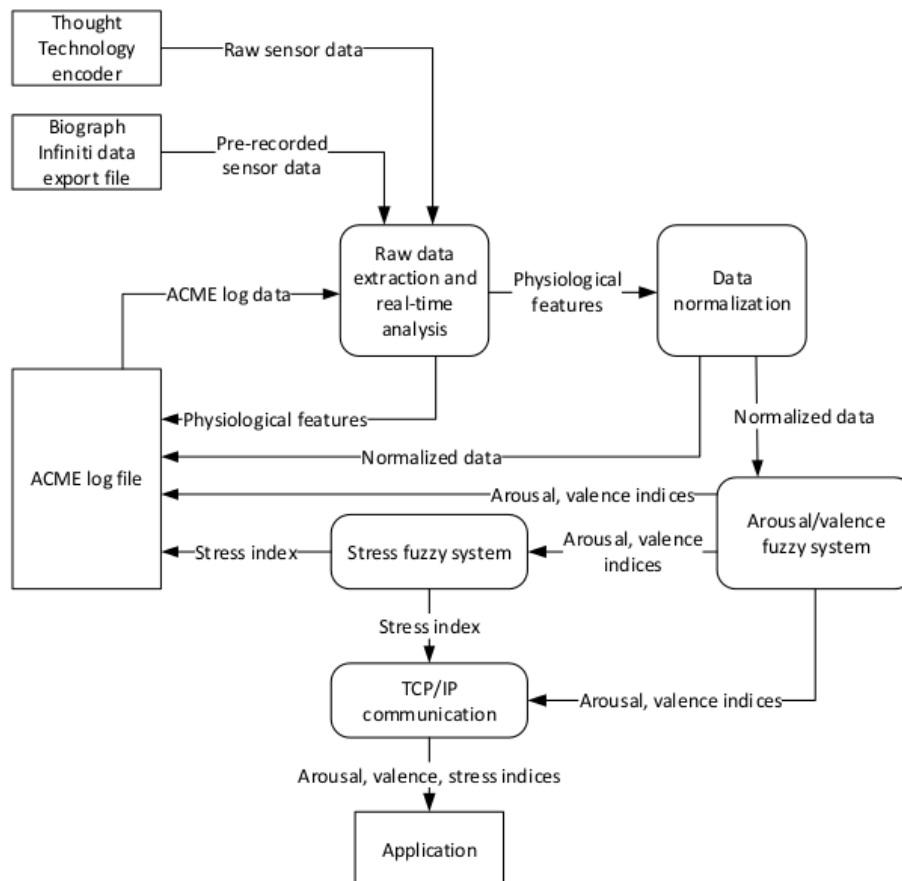


Figure 7.1: The ACME architecture

7.2.1 Input data

ACME is able to read raw physiological data from three different sources:

- A real-time data stream from one or more *Thought Technology* encoders;
- A text file which collects raw EMG, BVP and EDA data obtained by exporting physiological data previously recorded through the *Biograph Infiniti* software (Thought Technology, 2013);

- A log file generated by ACME during a previous session with the system. It collects raw sensor data as well as processed data.

The data stream collects the raw signals from all the physiological sensors connected to the encoders, at the maximum rate the sensors are capable of: EMG signals are sampled at 2048 Hz, while BVP and EDA signals are sampled at 256 Hz.

The text file exported by Biograph Infiniti collects raw EMG, BVP and EDA data recorded at 256 samples per second. ACME can detect the *markers* that Biograph Infiniti users can place on the raw data during the recording session, to point out a particular event occurred during the evaluation (e.g., a startle stimulus), or to signal the start and end times of a task carried out during data recording, such as the baseline. These markers are usually exported together with the physiological data, and presented as strings in correspondence to the relative timestamp. When one or more markers are detected, ACME allows users define one or more baseline sessions, by selecting one or more pairs of markers.

Finally, the ACME log data file is a text file that collects all the raw sensor data as well as the data processed by the system:

- Filtered physiological data;
- Mean values for each physiological signal (EMG from the zygomaticus and corrugator muscles, HR, EDA);
- Normalized values for each physiological signal (EMG from the zygomaticus and corrugator muscles, HR, EDA);
- Arousal and valence indices calculated by ACME;
- Stress index calculated by ACME;
- A flag indicating if ACME is recording a baseline.

7.2.2 Raw data extraction and real-time analysis module

This module can acquire raw physiological data from the three sources discussed in the previous section. For the second and third sources, ACME uses the data read from the text files to simulate a real-time signal reading. The module processes the data stream or the text files to generate multiple arrays of raw data sampled at 8 Hz, each corresponding to 125 ms of physiological data for a single signal. ACME is designed to read raw data from facial EMG (related to zygomaticus major and corrugator supercilii activity), BVP, and EDA.

Once the raw physiological data from the data reading module is acquired, the module filters it to reduce signal noise, caused for example by electrical interferences or user movements. To calculate the activity of zygomaticus major and corrugator supercilii, the module rectifies raw data using the RMS of the signal, then applies a low-pass filter using a 1 s moving window. For EDA, a low pass filter is applied to the raw data using a 5 s moving window, and the measure unit is changed from S (Siemens) to μS (microSiemens). BVP data is transformed using an adaptation of an open source peak detection algorithm (Zong et al., 2003), which allows to calculate IBI values and thus HR data. This module hosts also an implementation of a second peak detection algorithm, used to analyze ECG data, which is based on (Haag et al., 2004; Pan and Tompkins, 1985). The physiological features, as mentioned in the previous section, are memorized in the ACME log file.

7.2.3 Data normalization module

The data normalization module takes the physiological features extracted up to this point, and calculates mean and SD during baseline recording for each one. After the baseline calculation has completed, the module save mean and SD values and uses them to normalize the data sent by the previous module, to minimize physiological differences among users. The data is t -transformed so that all values lie in a 0–100 range, where 50 is the mean, and the extremes correspond to the mean value plus or minus 10 times the SD. The t -transformed data is memorized in the ACME log file.

7.2.4 Arousal/valence fuzzy system module and stress fuzzy system module

The arousal/valence fuzzy system module takes the normalized physiological values from the previous module, and translates them into indices of user's arousal and emotional valence, both in the 0–100 range. The fuzzy sets, the membership functions and the fuzzy rules are inherited from Mandryk and Atkins's original system. The fuzzy rules are the same as the ones proposed in the original system, because the authors grounded them in the theory of how physiology is related to the psychological concepts of arousal and valence, and also employed the circumplex model of affect (Section 1.1.2) to map stress in the arousal–valence space.

The membership functions were defined by Mandryk and Atkins through an analysis of a data set collected during the development. We initially modified the original functions because we consider 50 (in the 0–100 range) to be the mean value calculated during baseline recording (see Section 7.2.3). Conversely, in the system proposed by Mandryk and Atkins each physiological feature is normalized as a percentage of the span between the minimum and maximum values obtained for each user, which is an approach that assumes the use of a pre-recorded physiological data set. Furthermore, we decided to test the functions with physiological data recorded from users during a game session with a training application for evacuation in a fire emergency (Section 7.4.1). On the basis of our observations, we applied some slight modifications to the functions. For example, with respect to the original fuzzy sets defined for EMG, we widened the membership functions: this allowed ACME to reduce, in the valence index resulting from the fuzzy system, the width of fluctuations that we were able to observe during preliminary tests carried out for the development of the system. Figure 7.2 shows the fuzzy sets employed in the arousal/valence fuzzy system.

The stress fuzzy system module takes the arousal and valence indices and translate them into an index of user's stress. The stress index corresponds to the frustration index calculated in the original system (Mandryk and Atkins, 2007), which is defined as a high–arousal, low–valence affective state, and corresponds to the description of stress and anxiety in the circumplex model (Russell, 1980). Therefore, we did not modify neither the original fuzzy sets and membership functions, nor the fuzzy rules described in (Mandryk and Atkins, 2007). Figure 7.3 shows the fuzzy sets employed in the stress fuzzy system.

7.2.5 Communication module

The communication module creates and manages a client–server connection, and sends the stress index as well as the arousal and valence indices to the connected application, every 125 ms. A C# script has already been developed for Unity, which can be easily included in every project employing this game engine, that allow the application to automatically establish a connection with ACME and receive the arousal, valence, and stress indices. We tested the communication module with a Unity application to measure the latency introduced by the client–server connection. We used two PCs running the application and ACME respectively, connected through Ethernet to the university LAN. The results showed that the application received ACME data with a mean latency of about 1 ms.

7.3 ACME User Interface

Figure 7.4 shows the ACME user interface. The menu items on the top strip allow the user (i) to choose the desired input for ACME (see Section 7.2.1), (ii) to start and stop the server (which manages the raw data reading, and communicates with external applications through the communication module described in Section 7.2.5), and (iii) to define if HR is calculated from either the BVP (the default choice) or the ECG signal.

If ACME detects that the user has chosen either Biograph Infiniti export data file or an ACME log file that includes markers, it asks the user to optionally choose two separate markers among the available ones (including the start and the end of the file) to define a time interval during which the input data will be considered by ACME as data recorded during a baseline session.

Below the menu items, a message box keeps the user informed on the system state and various events, such as sensor connections, the server starting or stopping, or a successful connection from an external application.

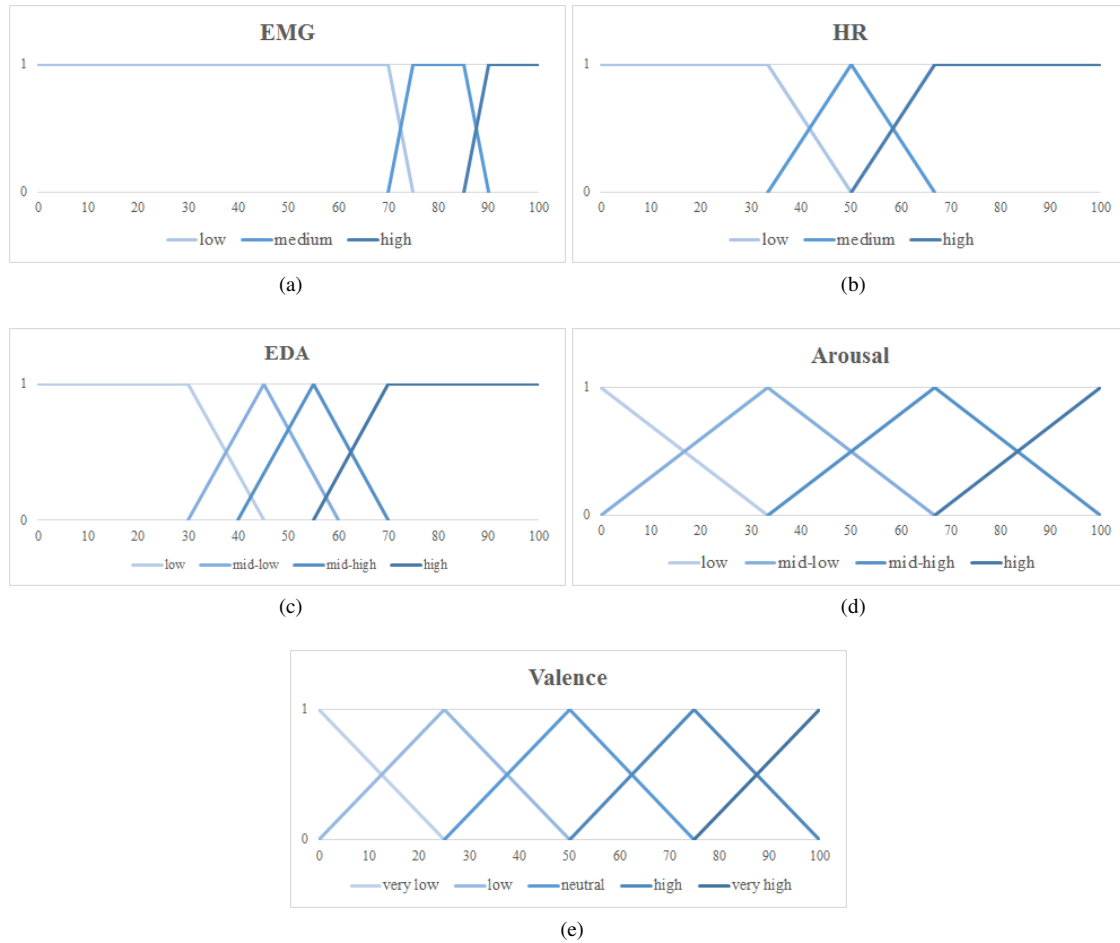


Figure 7.2: The fuzzy sets employed by ACME to translate normalized physiological data into arousal and valence indices.

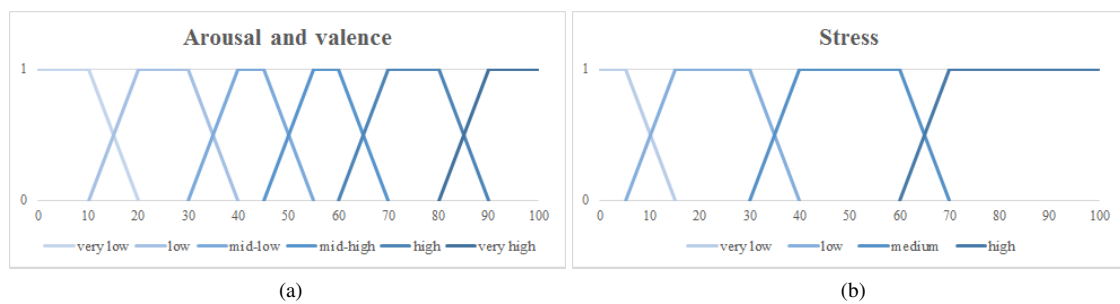


Figure 7.3: The fuzzy sets employed by ACME to translate arousal and valence indices into a stress index.

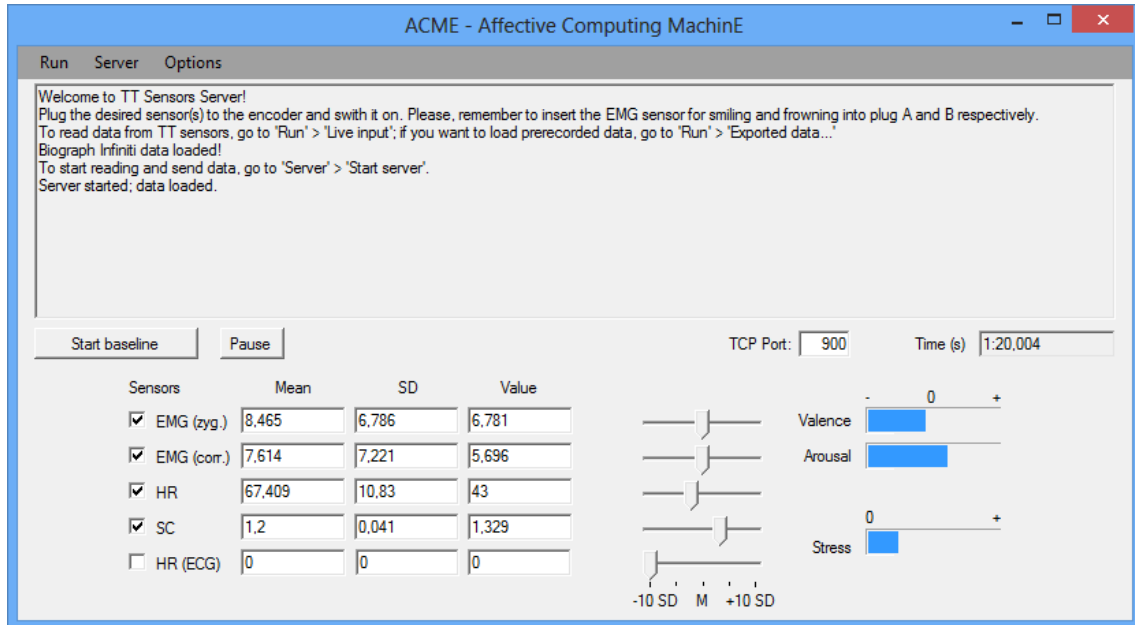


Figure 7.4: The ACME interface.

Below the text box, a button allows users to start or stop baseline recording, which is used to calculate mean and SD values for each physiological feature (Section 7.2.3). The baseline button allows users to manage the baseline recording session. The input data received by ACME before the start of the baseline recording is discarded, and the stream of data that follows the end of the baseline recording session is automatically considered by ACME as data to be normalized using the mean and SD values calculated from the baseline data. A second button allows users to pause the analysis of physiological data. When on on pause state, ACME discards the raw data if it is gathered from physiological sensors; otherwise, if the data is gathered from an export file or from an ACME log, the data extraction and real-time analysis is paused.

An interactive text box under the message box allows to define the TCP port used by the communication module (Section 7.2.5) to send data to external applications. A second text box shows the time, in seconds, elapsed from the server start.

On the bottom half of the application window, ACME shows (from left to right) (i) all the accepted physiological signals (ticked if the relative sensor is connected), (ii) the mean and SD values calculated during the baseline recording, (iii) the current value read from input, (iv) the indices representing the normalized physiological values from -10 SD to $+10\text{ SD}$ (Section 7.2.3), (v) the valence, arousal, and stress indices. Note that valence and arousal indices can be negative or positive, while the stress index can be only positive (0: not stressed; 100: completely stressed).

7.4 ACME Preliminary Evaluation

For a preliminary evaluation of ACME, we carried out a study using two data sets: (i) the physiological data set recorded during a pilot test designed to elicit states of high arousal and stress in users by immersing them in a stressful VE, and (ii) the physiological data set recorded during the use of the mobile breathing training application described in Chapter 6. We have chosen these two data sets because we wanted to compare the physiological responses to experiences that would elicit different affective states: as observed by Ward and Marsden (2004), significant differences in bodily responses may be evident only in the case of major differences in the experience.

In the following, we describe the pilot study in the stressful VE, as well as report the results of the comparison analysis of the data.

7.4.1 Pilot study with a training application for evacuation during a fire emergency

In this study, EDA, HR, and facial EMG data were recorded to collect the physiological responses of participants immersed in a training simulation, developed to teach correct behaviors during mass emergencies. The simulation depicts, with immersive 3D graphics, a mass emergency involving a number of potentially stressful stimuli, such as sudden and loud explosions, fire and smoke, the sight of wounded people asking for help, and blocked evacuation routes (Figure 7.5). The application employs physics to realistically simulate not only effects like smoke, but also the effects of explosions on virtual humans and train station structure: for example, an explosion can throw away nearby virtual humans, and can cause the collapse of a platform roof.

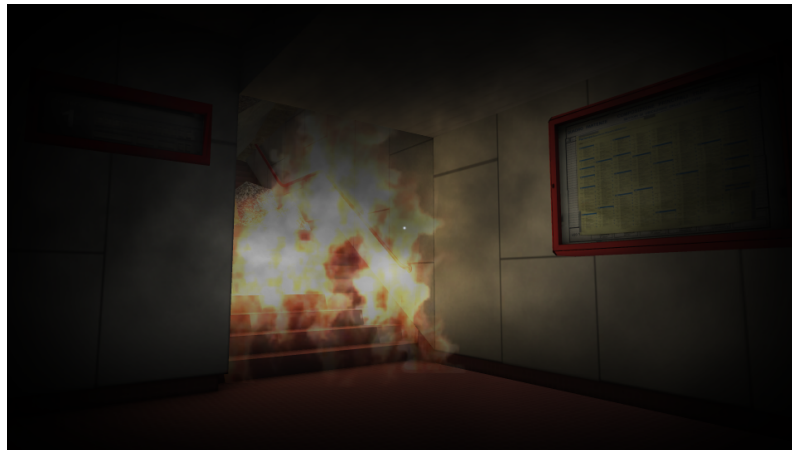


Figure 7.5: One of the stressful stimuli: fire is blocking one of the escape routes.

Users have to evacuate the station without being hurt, approaching correctly each possible source of risk: for example, in case of smoke, they have to crouch as fast as possible to avoid breathing it. The focus of this pilot study was to evaluate participants' physiological responses to the mentioned stressful stimuli, which were accompanied by visual and auditory cues: for example, explosions were accompanied by a loud, low-frequency sound, and the shaking of the on-screen image; remaining inside a cloud of smoke would cause the reduction of the field of view, as well as intense coughing, followed by death in a few seconds; touching fire or metal debris would cause the screen to flash in red, and the reproduction of shouting sounds. The volume of explosion noises is significantly higher than the background noises (ambulance sirens, screams, and collapsing debris) to startle the user. Furthermore, if an explosion happens sufficiently close to the user, a high-frequency buzz is played for a few seconds to simulate ear ringing, and red bars are shown around the screen to simulate a reduction of users' field of view.

At the beginning of the simulation, the user is located outside the train station. She has to cross the street and reach the train station, where she has to buy a ticket at an electronic ticket booth. Then, the user has to take the underpass and reach the platform where her train is waiting. Once the user has got on the train, she has to find her place (pointed by a checkpoint) and sit here. During this first phase of the simulation, written messages showed on screen and checkpoints are employed to help the user. Realistic train sounds and annunciations were played to increase user immersion.

After few seconds, a tank wagon explodes in the station (Figure 7.6). The explosion hits all the trains in the station as well as the user's wagon, which causes her to pass out. When the user wakes up, a message tells her to flee to safety.

The user has to evacuate the train station by finding a safe path through the corpses, wounded people, shattered wagons, and debris. The VE was designed to provide users with a predefined safe path, shown in Figure 7.7, which was long and difficult enough to allow only one participant to complete it within the 10 min limit. However, because the animation of the collapsing debris is calculated in real-time by the physics engine employed by the simulation, explosions sometimes can open alternative paths. For example, some parts of platform roofs can fall and lean on wagons left intact by the explosions, allowing users to climb



Figure 7.6: The moment when the user, sat on the wagon, see the explosion.

over these wagons and save themselves.

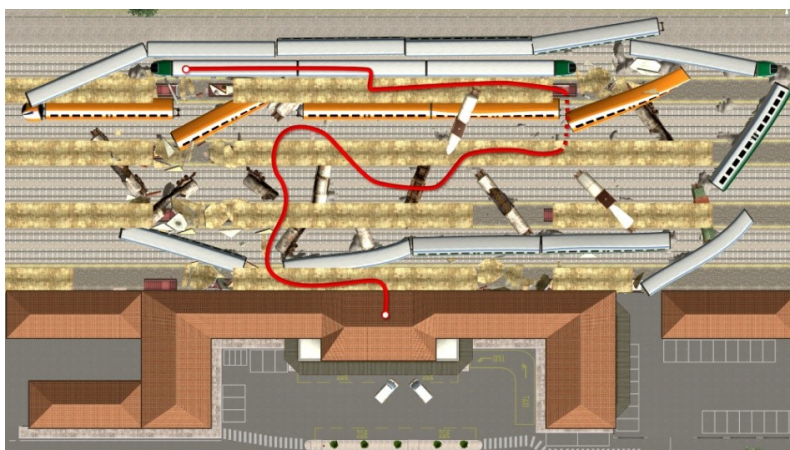


Figure 7.7: The predefined safe path, shown as a red line over a top view of the train station. The dotted portion of the line indicates that the user has to take the subway.

10 participants took part to the pilot study. For each participant, we recorded left corrugator supercillii EMG, zygomaticus major EMG, EDA, and HR. Before carrying out a 10-min play session with the training application, participants practiced the game controls in a training session, during which they navigated inside a 3D VE along a path designed to make them use each possible control (walking, running, crouching, ...) to reach a sequence of checkpoints (Figure 7.8). Participants' baseline was recorded during the training session.

A qualitative analysis of physiological data showed that the experience in the VE was able to increase participants' stress level: the considered stressful stimuli were able to generate intense SCRs, and that the activity of the left corrugator supercillii increased gradually during the evacuation, in particular when the participants found that the chosen escape route was blocked.

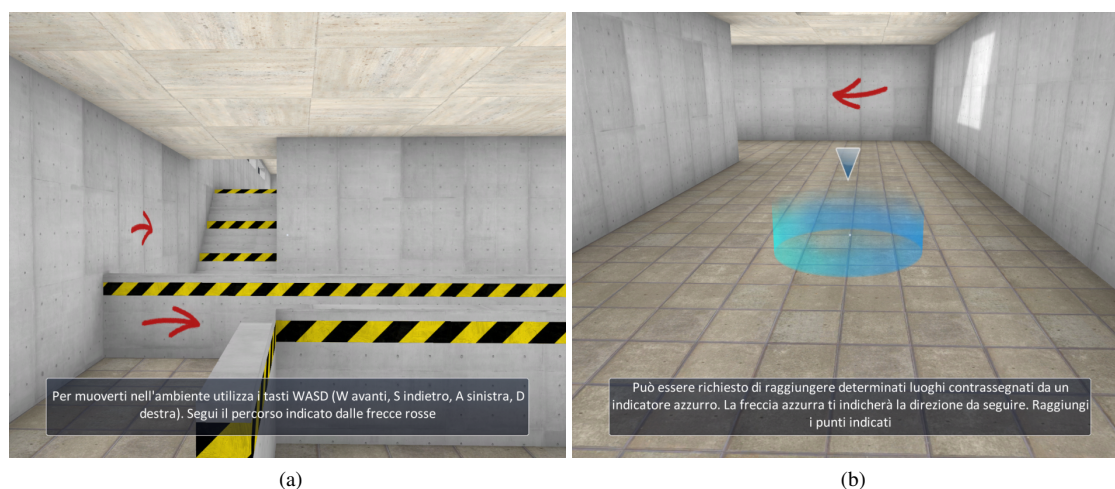


Figure 7.8: Two screens of the 3D VE employed for the training session. (a) A maze designed to train users to employ W, A, S, and D keyboard keys to move inside the 3D VE; (b) a checkpoint along the training path.

During the study, for each play session we recorded the PC screen using a webcam. By analyzing the recorded videos and the ACME log data files (see an example in Figure 7.9), we were able to evaluate the ability of ACME in effectively detecting any possible physiological responses of participants to specific stimuli such as the explosions.

Since we were able to employ ACME log files as input to ACME, we were able to test offline the effects of changes in the two fuzzy logic system employed by ACME (Section 7.2.4) and thus improve the ability of the system in correctly detecting stressful responses in participants, and at the same time reduce false positive errors, i.e., erroneous detection of stressful responses caused, for example, by artifacts in one or more physiological signals.

7.4.2 Preliminary evaluation of ACME

The physiological data collected during the study described in Section 7.4.1 was compared to the ACME output resulting from the physiological data of the first 10 participants who took part in the evaluation described in Chapter 6. Only the arousal index was computed; stress values were not calculated because one of the two data sets do not employ EMG recordings, essential for the calculation of valence and thus to discern from stress and excitement.

The goal of the preliminary evaluation described in this section was to verify that ACME was at least able to discern correctly between the physiological responses elicited by two different experiences: a play session with a VE-based training application designed to immerse the user into a stressful situation, and a session with a mobile breathing training application meant for stress and anxiety treatment (Chapter 6).

First, we employed ACME to calculate participants' arousal index from the physiological data recorded during the three conditions described in Section 6.4. The analysis reported that the Wave condition elicited the lowest level of arousal. Then, we compared with a *t*-test the arousal level elicited by the Wave design in the breathing training mobile app with the same index elicited by the virtual experience described in the previous section. The result seems to confirm that ACME can successfully discern between the arousal level elicited by a stressful experience inside a VE and the breathing exercise ($t = 2.63$, $p < 0.05$). Figure 7.10 shows the mean arousal index values in the four conditions.

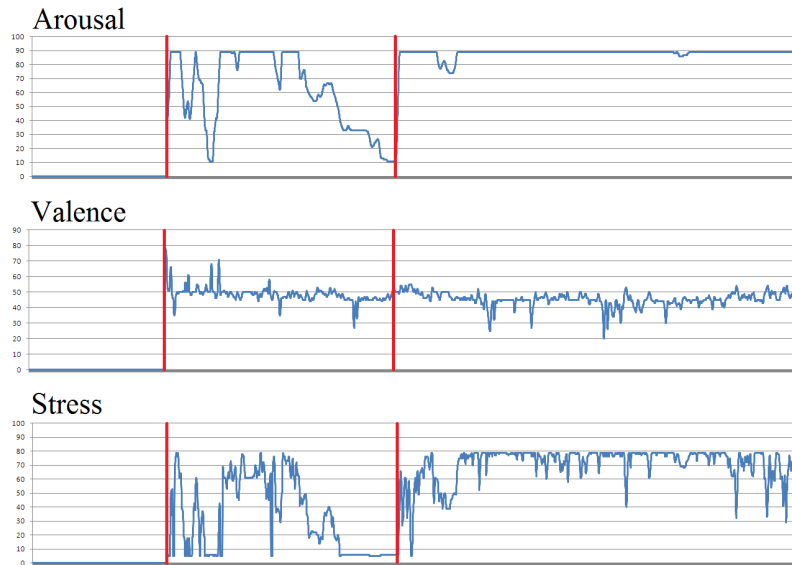


Figure 7.9: Arousal, valence, and stress indices recorded from the Participant no. 2 during the experiment. The first red marker points out the end of the baseline session; the second red marker points out the moment when the first explosion occurs during the play session, i.e., when the participant was sit on the wagon. Note how the arousal and stress indices sharply increase in correspondence of the second marker.

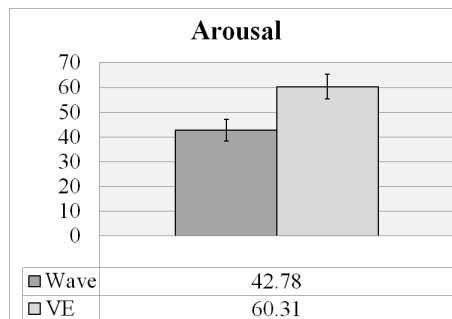


Figure 7.10: Mean values of arousal index data. Error bars indicate standard error of the mean.

8

Physiology–Based Evaluation of a Biofeedback–Controlled Game for Relaxation Training

In Chapter 6, we started to explore the use of physiology to evaluate the effectiveness of a mobile breathing training application which, however, does not integrate ACS functionalities, and thus could not change its state in response to variations in users' breathing pattern or stress level. In this chapter, we add the missing link focusing our attention on the use of biofeedback (see Section 2.3.1) in the field of relaxation training, which is an application of affective computing with important implications for health and well-being (Chittaro and Sioni, 2014). For this reason, we propose a novel application for relaxation training that combines ideas from affective computing and games. A version of the application employs ACME, the ACS proposed in Chapter 7, to detect user's level of stress, using it to influence the affective state and the behavior of a virtual character in a realistic 3D VE. In the study presented in this chapter, we evaluate the effectiveness of this version of the game by comparing it with a less costly (but potentially less accurate) approach that uses a single physiological signal (EDA). To improve the traditional comparison methodology employed in affective computing studies, we add a placebo condition in which user's stress level, unbeknown to him/her, is determined pseudo-randomly instead of taking into account his/her physiological sensor readings. The results of the evaluation show that only the feedback presented in the EDA-only version of the application is perceived as significantly more accurate than the placebo condition. If only the two non-placebo conditions had been considered, their effectiveness would have instead appeared similar. This outcome highlights the importance of using more thorough methodologies in future affective computing studies.

8.1 Introduction

Many studies in the biofeedback literature still employ very simple acoustic or visual cues to provide users with feedback about their stress level, e.g., alarm sounds that are activated when a physiological signal reaches a certain threshold, or virtual thermometers whose level is directly related to users' arousal. However, in recent years, some studies started to employ richer stimuli such as virtual reality and video games to convey information to users (e.g., Bersak et al., 2001; Bouchard et al., 2012). An advantage of this approach is that immersive and realistic VEs draw more of the users' attention and make the feedback more effective (Bersak et al., 2001).

The goal of this chapter is threefold. First, we propose a novel biofeedback application for relaxation training that combines ideas from affective computing and games. By exploiting realistic 3D graphics, we present users with real-life simulated scenarios in which they have to focus on maintaining relaxation when faced with stressors. Players see a virtual character in third-person view, and the level of stress detected from their physiological signals is used to determine the facial expressions as well as the behavior of the virtual character. This relation between player's affective state and virtual character behavior provides a compelling visual and acoustic feedback to players about their current state of relaxation (Section 2.3.1).

To detect users' level of stress, the game can employ up to four physiological sensors: EDA, HR, and two EMG sensors, which respectively measure the activity of corrugator supercilii and zygomaticus major muscles.

Second, we compare the perceived quality of the feedback provided by two versions of the game, which differ only in the stress detection algorithm used. One version of the game employs ACME, the ACS proposed in Chapter 7, which exploits all four physiological signals; the second version of the game uses an algorithm inspired by a different approach in the affective computing literature, and exploit only the EDA sensor. We are interested in comparing these two approaches in the context of biofeedback–based relaxation training because in general they may have different advantages: a single–sensor approach is more practical and less costly; however, one can also hypothesize that the use of multiple physiological sensors may improve the accuracy of arousal and stress detection, because multiple signals allow one to consider different bodily responses at the same time (see Section 2.2).

Third, we aim to improve the traditional comparison methodology employed in affective computing studies, by adding a placebo condition in which users' stress level is determined pseudo–randomly instead of taking into account his/her physiological sensor readings. In medical studies, placebo conditions (i.e., conditions in which a sham instead of a real treatment is administered to participants) are commonly employed because factors such as participants' suggestibility could lead to improvements even with sham treatments. The purpose of the placebo condition is therefore to experimentally evaluate if the proposed treatment (e.g., a pill containing a new drug) is really superior to the sham treatment (e.g., a pill that looks like the real drug but does not contain it).

In affective computing studies, one cannot rule out the possibility that the suggestibility of participants could alter their perception of the application feedback, leading them to believe that a real biofeedback system is in place even if sham biofeedback is actually provided. The use of a placebo condition is common in the biofeedback literature for the evaluation of the long–term efficacy of biofeedback therapy based on simple stimuli, often with interesting results (e.g., Hunyor et al., 1997). In studies of ACSs, it could be interesting to include placebo conditions in long–term evaluations as well as in evaluations of the immediate perceived accuracy of the rich feedback provided by the application to determine if the—often complex and costly—affective computing techniques are actually playing a significant role in the effectiveness of the feedback. However, to the best of our knowledge, this has not yet been explored in the affective computing literature.

The chapter is organized as follows. In Section 8.2, we review the literature on biofeedback applications and evaluation of ACSs. Section 8.3 describes in detail the proposed biofeedback serious game. Then, Sections 8.4 and 8.5 describe in detail the experiment and its results, discussed in Section 8.6.

8.2 Related Work

Biofeedback systems for stress treatment mostly provide very basic feedback to the user. For example, in (Foster, 2004; Jensen et al., 2009) a repeating tone was employed, which decreased in frequency as the EMG activity of masseter and frontalis muscles respectively decreased, indicating a lower level of arousal and stress (Jensen et al., 2009), or a diminishing intensity of the disorder under treatment (Foster, 2004). The very basic feedback provided by other systems is visual, e.g., showing the signal graph of one or more physiological signals like frontal EMG (Fentress et al., 1986) or a thermometer whose temperature is positively correlated to EDA level so that higher arousal is represented by a higher temperature (Critchley et al., 2001, 2002).

The use of much richer feedback based on realistic and immersive VEs has been suggested in biofeedback training for relaxation and stress management, because VEs draw more of the user's attention, providing a much more appealing and effective feedback for achieving the correct affective state (Bersak et al., 2001). The biofeedback system described in (Bersak et al., 2001) is a racing game in which increasing relaxation level helps the user move faster inside the game, and the player must cope with the high level of arousal elicited by the competitive environment of the racing experience. In military applications, graphically intense games which employ biofeedback have been suggested for increasing soldiers' mental resilience and the effectiveness of their stress management skills. In particular, Bouchard et al. (2012) cast doubts on the efficacy of classic approaches to stress management training in the military, which are essen-

tially based on providing information and teaching techniques in a classroom, and highlight the importance of practice. For this reason, they added biofeedback-based audio and visual cues to stressful commercial video games as a mean to show increasing arousal (derived from HR and EDA signals). The cues included a decreasing field of view or the sound of a heart beating at an increasing frequency and loudness. They report that this game-based approach was effective in increasing the effectiveness of stress management. Similar feedback cues were employed by Dekker and Champion (2007), but for entertainment purposes, enhancing game play by detecting players' arousal calculated from HRV and EDA. They report gameplay issues that may have diminished the biofeedback effects in their study: usability issues caused by wearing physiological sensors, different involvement in game play shown by novice and experienced players, and different factors (such as room illumination) that had a negative effect on participants' concentration and engagement. Kato (2010) reviews biofeedback-controlled video games used in hospital settings to treat patients' with pediatric voiding dysfunction or *irritable bowel syndrome* (IBS), and were able to improve the treatment outcome thanks to the ability of game interfaces to increase interest and motivation, engaging the patient in the therapy.

The biofeedback literature provides various examples of comparisons between real and placebo biofeedback conditions (e.g., Hunyor et al., 1997; Rupert and Holmes, 1978; Tsai et al., 2007), which often led to relevant results: for example, Hunyor et al. (1997) observed that their biofeedback treatment of hypertension is successful in lowering systolic blood pressure with both a real and a placebo feedback signal.

In the affective computing literature, to the best of our knowledge no comparisons with a placebo condition are performed (see Section 2.2) to assess the effectiveness of ACSs. This is acknowledged in (Picard and Goodwin, 2008) which explicitly advocates the consideration of the placebo effect in experimental designs that evaluate the efficacy of ACSs.

8.3 The Biofeedback Game

By combining ideas from affective computing and games, we propose a novel application for relaxation training based on biofeedback. The proposed game exploits realistic 3D graphics to present users with real-life simulated scenarios in which they have to focus on maintaining relaxation when faced with stressors.

Various studies in the literature show that treatments for stress-related disorders that employ realistic VEs are as effective as (and in some case better than) traditional cognitive-behavioral therapies (e.g., Baños et al., 2011; Gorini and Riva, 2008; Villani and Riva, 2012). Unlike many biofeedback systems that employ very simple feedback (see Section 8.2), we focused on using 3D graphics to provide players with embodied feedback through a virtual character. Once the system infers in real-time the player's stress level from his/her physiological signals (EDA, HR and EMG from corrugator supercilii and zygomaticus mayor muscles), the virtual character in the VE reflects the player's stress level through facial expressions, body postures and movements. The emotions expressed by the virtual character through its facial expressions range from happiness to anger, which are basic emotions that can be universally associated to unique facial expressions (Ekman and Friesen, 1971). The use of these facial expressions in the proposed game should therefore make it easy for players to understand the emotions conveyed by the virtual character.

The sense of embodiment, i.e., the connection between the player's body in the physical world and the body of the virtual character which represents the player in the VE, is considered to be a key element for the efficacy of VE-based relaxation treatments (Pallavicini et al., 2013; Riva et al., 2007; Villani et al., 2007). In particular, the more intense is the connection, the greater is the player's sense of being present inside the VE (Riva and Mantovani, 2012).

Each level of the biofeedback game shows the virtual character carrying out a task inside a realistic VE, such as an office, a school building or a train station. Inside each VE, various stressors can affect the user (and the virtual character that displays his/her affective state). To play the various levels of the biofeedback game, users have to relax and remain centered on their relaxation state, without being distracted or affected by the stressors. By doing so, they allow the virtual character to accomplish goals in the VE. For example, in one of the levels set in the school building VE, the goal is to evacuate safely during a fire emergency; the stressors are a combination of alarm sound, explosion sound and visual cues (smoke clouds and injured persons). The virtual character gets slower and slower as the player's level of stress increases, eventually coming to a halt, shivering, until the player is able to reach a sufficient level of relaxation that allows it

to resume walking. In another level set in the school building, the player is attending a class and has to attentively listen to the lecturer, but the other students around him/her produce annoying chatter at three different times during the brief lecture; if the player’s level of stress increases, the volume of the lecturer’s voice decreases, making it harder to achieve the goal. The use of stressors is currently employed by some of the biofeedback–based applications for relaxation training proposed in the literature (e.g., Bersak et al., 2001; Bouchard et al., 2012): by exposing players to stressful stimuli, the player not only learns how to relax, but (s)he learns to do it in a stressful environment, which is a fundamental element of biofeedback–based therapies such as SIT (see Section 2.3.1).

8.3.1 First level of the game

In the first level of the game, which is the one employed in the experiment described in this chapter, the virtual character is called Andrew and the introductory screen informs the player that Andrew has just been hired by a big company and put to work on writing a business plan. The level shows Andrew working in front of his PC, inside a typical office environment.

The goal of the level is to allow Andrew to complete the assigned task; the more the user is relaxed, the more the character is relaxed and the larger the final score (percentage of work completed by Andrew) will be at the end of the 3 minutes the level lasts. To reach the goal, Andrew has to remain relaxed and work continuously at the PC, ignoring a distracting stressor, i.e., a phone on its desk that occasionally rings. In particular, we have set this stressor to occur for 3 times: the first time for 10 s, the second for 20 s, and the third for 30 s. The first, second and third occurrence of the stimulus respectively start 20 s, 50 s, and 90 s after the beginning of the level. No other stressors are employed in this level of the game. The setting and the task we designed for this level of the game was chosen because it is a common situation in an office context, and thus should be easy to understand for players.

To play the level, the user does not interact with Andrew using traditional game controllers such as mouse, keyboard, or gamepads; but by trying to control his/her relaxation state and to ignore the ringing phone. If the user is calm and relaxed, then Andrew is focused on its work, showing a smile on its face. As the users’ stress level increases, Andrew is more and more unsettled, less and less focused on the work, and shows discontentment through facial expressions and mumblings of growing intensity.

During phone ringing, the virtual camera slightly zooms out to include in the current view the phone placed over its desk.

Andrew can be in one of five different states, each one characterized by distinct behaviors (i.e., body gestures, facial expressions and vocal expressions), depending on user’s level of stress. In the following, we describe in detail all behaviors employed in the first level of the game, listing the states of the character in increasing level of stress:

- *State 1*: the character is completely relaxed. Seated on its chair, Andrew is working at the PC using its mouse and keyboard (Figure 8.1a and 8.1b), and watching the PC screen with a smile on its face (Figure 8.1c);
- *State 2*: the character starts to show signs of uneasiness. Andrew continues its task (Figure 8.1d and 8.1e), but sometimes moves its chest towards the screen (Figure 8.1d), mumbles, and shakes its head; it starts to frown, and its smile starts to fade (Figure 8.1e);
- *State 3*: the character is more unsettled than State 2. Its head shakes more frequently, and the mumblings are louder; frowning is more visible, and the smile has disappeared (Figure 8.1g). Andrew tries to continue working as in Figure 8.1a and 8.1b, but sometimes stops for a few seconds, reclines back and puts its hands on the legs (Figure 8.1f);
- *State 4*: the character cannot focus on its task anymore. Andrew stops working (Figure 8.1h); the mumblings get increasingly loud, and its face becomes very tense and angry (Figure 8.1i). Sometimes, the character raises its hands (Figure 8.2a and 8.2b), and throws a punch on the table (Figure 8.2c);

- *State 5*: the character reaches maximum level of stress. Its face becomes angrier (figure 8.1k), it raises its hands (Figure 8.1j) and throws a sequence of punches on the table (Figures 8.2b and 8.2c); its mumblings turn into shouts.

The animations employed in the five states were built by performing multiple sessions of human motion capture with an 8–cameras OptiTrack ARENA system, to make Andrew’s body movements as realistic as possible. Facial expressions were modeled by hand, and were intentionally slightly exaggerated to make it easier for users to interpret them. Facial and body animations were then integrated in the game in such a way that, during a behavior change, the first behavioral animations can seamlessly blend into the second one, without abrupt changes that would make the whole behavioral sequence unrealistic.

We defined the five states in such a way that the stress level displayed by Andrew could grow progressively from the first to the fifth. To pre–test the understandability and the clarity of the five states and evaluate if users could correctly interpret them, we conducted a pilot study with 10 participants (7 M, 3 F), with a mean age of 31 (SD = 4.08). We recorded a 10–sec video for each one of Andrew’s states, and showed the five videos to each participant following a latin square design. Then, we asked participants to rank the videos from the one showing Andrew at the lowest level of stress, to the one showing Andrew at the highest level of stress. All participants were able to correctly rank the five states without making any errors.

8.3.2 Users’ stress detection

To detect users’ stress, our study tests two different algorithms. The first employs only the EDA signal, and derive from the algorithm presented in (Healey and Picard, 1998); the second is implemented in the ACME ACS described in Chapter 7. We have chosen these two well–known algorithms because they are more generally representative of two categories of stress detection algorithms: a less costly approach that uses a single sensor (EDA), and a potentially more accurate approach that adds three more sensors (HR, and EMG of corrugator supercilii and zygomaticus major). The EDA–based algorithm by Healey and Picard was originally developed to allow detection of users’ SCRs through a wearable computer, while the ACS presented in (Mandryk and Atkins, 2007), which ACME is based on, was developed to detect not only users’ arousal and valence state, but also the level of five discrete emotions: fun, challenge, boredom, frustration, and excitement.

In implementing the EDA–only algorithm, we generally followed the approach described in (Healey and Picard, 1998) but we changed the filtering of the EDA signal to reduce latency as much as possible. We keep using the first forward difference of the signal (as in the original algorithm) to analyze the signal slope, but for interpreting it as a trend of player’s stress level instead of detecting SCRs.

In the following, we describe how the two implemented algorithms compute a user’s stress index from the current user’s physiological state, and how the stress index is instead generated pseudo–randomly in the placebo condition of our study.

The EDA–only algorithm

As in the original algorithm by Healey and Picard (1998), our implementation of the EDA–only algorithm does not derive users’ stress from the absolute EDA value, but from the signal shape. At the beginning of a session, the stress index is set to zero and this makes the virtual character start with its lowest stress state (State 1). Then, the algorithm identifies rising and declining intervals of the EDA signal (when the user is confronted with a stressor, the EDA signal typically rises fast and then declines slowly) by sampling the EDA signal every 0.5 s. The algorithm does not use a moving window for signal sampling, but keeps and updates a temporary baseline value as described in the following:

- If the EDA starts increasing (i.e., the current EDA value is higher than the value of the previous sample), the EDA value at the onset is memorized as a temporary baseline value;
- During a signal increasing phase, when the EDA exceeds for the first time a $0.05 \mu S$ threshold (recommended in (Boucsein, 2006) to distinguish a skin conductance response from signal fluctuations) over the temporary baseline value, the stress index is incremented by 1. If the EDA signal keeps

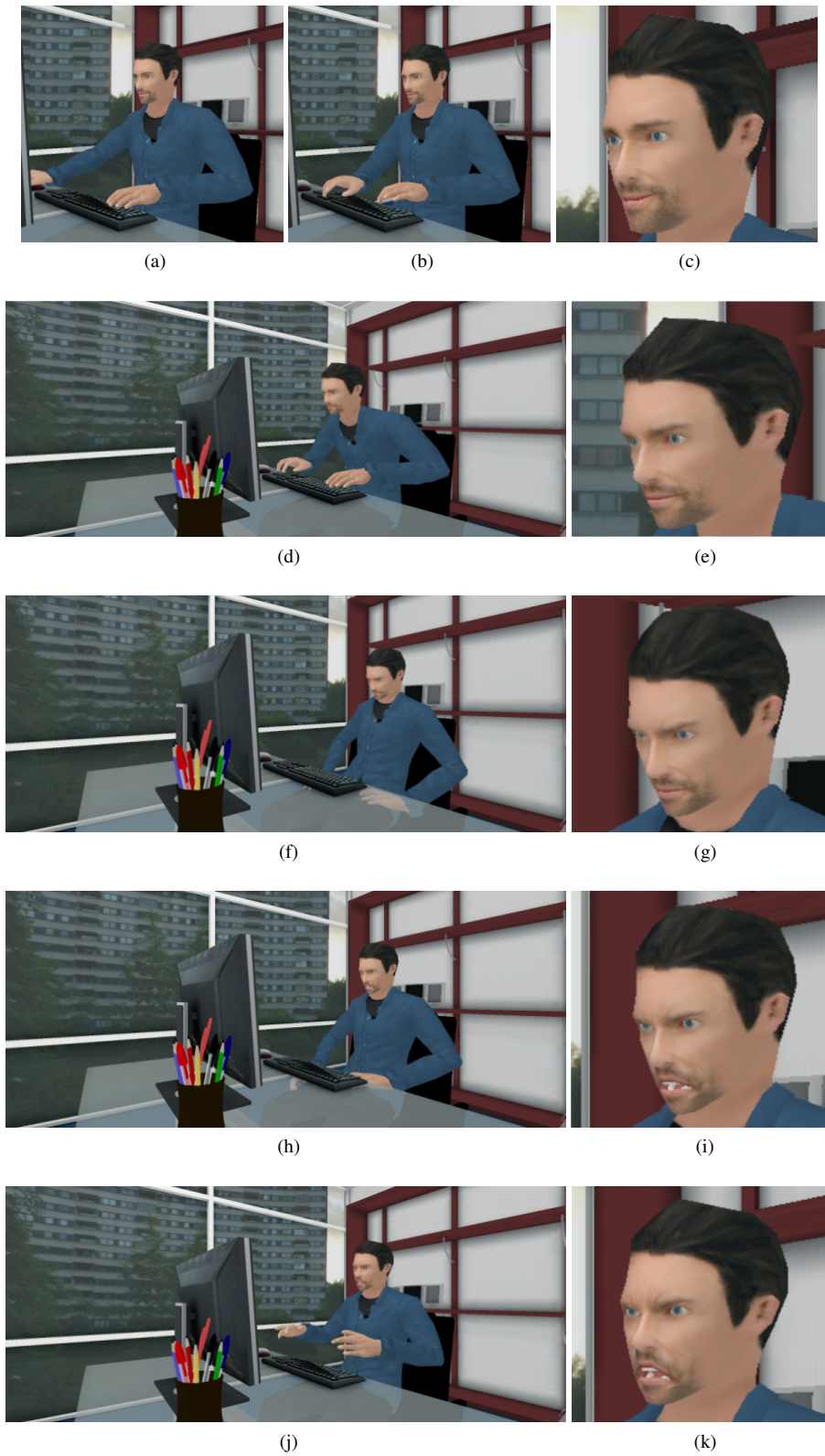


Figure 8.1: Screenshots of the virtual character in its five different states.

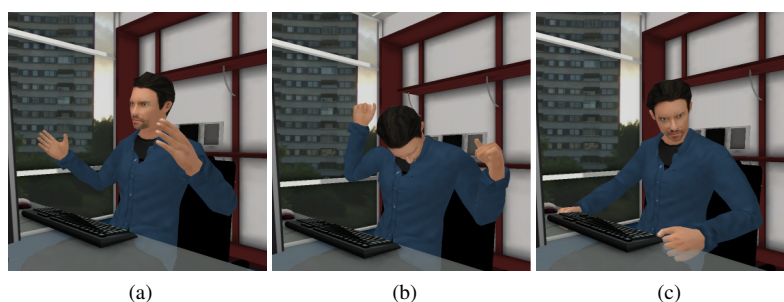


Figure 8.2: Screenshots of the virtual character raising its hands and punching the table.

increasing, the stress index (and consequently the virtual character state) is incremented by 1 every 3 s. If the stress index is 4, no further increment is applied. From an increasing EDA, therefore, we generally infer an increasing stress level in users.

- When the EDA starts decreasing (i.e., the current EDA value is lower than the value of the previous sample), the current EDA value is memorized as the temporary baseline value, the stress index (and consequently the virtual character state) is decremented by 1 every 3 s. If the stress index is zero, no further decrement is applied. From a decreasing EDA, therefore, we generally infer a decreasing stress level in users.

Figure 8.3 shows an example of use of the EDA-only algorithm.

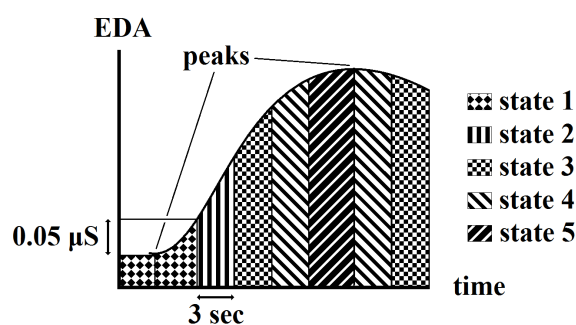


Figure 8.3: An example of use of the EDA-only algorithm.

This algorithm is simple to implement, and is designed for real-time signal analysis. However, it employs only a single physiological signal to infer user's stress. It does not need a baseline value measured in advance by recording the physiological data of players in a state of maximum relaxation, which is the classic approach in the psychophysiology literature. Instead, it employs a temporary baseline value that is continuously updated during the game session to follow the player's trends of increasing and decreasing relaxation. Finally, it must be noted that, as in (Healey and Picard, 1998), the algorithm does not decompose the EDA signal into phasic and tonic components (see Section 1.3.1), since this operation requires filtering operations that may introduce latency in the subsequent analysis of the SCL component. Therefore, the EDA-only algorithm analyzes the EDA signal as a whole.

The multi-sensor algorithm

The multi-sensor algorithm exploits ACME to calculate users' stress level. Unlike the previously described EDA-only algorithm, which focuses on the trend of the EDA signal, the multi-sensor algorithm focuses on the actual value of the EDA recorded by the electrodes. Like the EDA-only algorithm, the multi-sensor algorithm does not decompose the EDA signal into phasic and tonic components (see Section 8.3.2).

Once the stress index is computed by ACME as described in Chapter 7, we map its values into the state of the virtual character as follows: (i) 0–55 range: state 1, (ii) 55–75 range: state 2, (iii) 75–86 range: state 3, (iv) 86–94 range: state 4, (v) 94–100 range: state 5.

Unlike the EDA-only algorithm, this multi-sensor approach should be able to distinguish negative stress from positive arousal, and requires a baseline recording session to normalize data.

8.3.3 The placebo condition algorithm

In the placebo condition, the state of the virtual character is assigned by the game without taking into account user's current physiological state. A completely random assignment would make the behavior of the character look implausible and unrealistic, so we resorted to a pseudo-random approach for creating sham biofeedback as follows: (i) we take a physiological data recording of a previous game session (ii) we pick one of the two previously described stress detection algorithms at random, (iii) we feed the pre-recorded physiological data into the algorithm, starting from a random second (in the 0–179 range) of the recording, (iv) if the sham data stream ends while the user is still playing the game, we restart the stream from second 0 of the recording.

8.4 Experimental Evaluation

The evaluation of the proposed biofeedback game followed a within-subject design, with *stress detection technique (EDA-only, multi-sensor, placebo)* as the independent variable.

8.4.1 Participants

The evaluation involved a sample of 35 participants (26 M, 9 F) recruited through direct contact among graduate and undergraduate students at our university and people from other occupations. Participants were volunteers who received no compensation.

8.4.2 Materials

The game was run on a PC, and displayed in full-screen mode on a 30", 2560 × 1600 pixel LCD monitor. The distance between the screen and the user was about 1.5 m. The lights in the room used for the evaluation were turned off to prevent issues in gameplay experience reported in the study by Dekker and Champion (2007). We also employed a much larger display for the same reason.

Participants' physiological data was recorded and processed on a second PC. To record participants' physiological data, we employed five sensors, following the placement suggestions described in (Andreassi, 2007):

- Two EMG sensors coupled with disposable pre-gelled electrodes, placed over the zygomaticus major and the corrugator supercilii muscles;
- An EDA sensor placed on the intermediate phalanges on the middle and ring fingers;
- A PPG sensor for BVP, placed over the distal phalanx of the index finger.

The physiological data was recorded at 2048 Hz with a Procomp Infinity encoder, and processed in real-time by the algorithms described in Sections 8.3.2 and 8.3.2. The calculated stress index was sent to the PC running the biofeedback game through a LAN connection. We assessed the latency of the LAN connection: the time required to receive the stress index in the game was about 1 ms, which is too short to be perceivable by participants. To record participants' subjective opinions about the biofeedback game, we employed a questionnaire that assessed:

- *Perceived quality of the biofeedback.* To assess this aspect, after each condition users rated three items ("The character's relaxation level corresponded to mine", "When I was relaxed, the character was relaxed too", and "When my relaxation level was changing, the character's relaxation level

immediately changed too”) on a 7–point Likert scale (1 = strongly disagree, 7 = strongly agree). Participants’ ratings of the three items were averaged to form a reliable scale (Cronbach’s alpha = .72);

- *Difficulty of relaxation training.* To assess this aspect, after each condition users rated three items (“I found it difficult to increase my relaxation level”, “I had to struggle a lot to increase my relaxation level”, “The game allowed me to easily increase my relaxation level”) on a 7–point Likert scale (1 = strongly disagree, 7 = strongly agree). Since the first two items contained negative statements and the third item contained a positive statement, we reversed the score of the third item, then participants’ ratings of the three items were averaged to form a reliable scale (Cronbach’s alpha = .77).

8.4.3 Procedure

Participants were verbally briefed about the nature of the task, the data measured by the physiological sensors, and the anonymity of the collected experimental data. They were asked to try the three versions of the game level described in Section 8.3.2. They were told that, in all three versions of the level, their goal was just to relax as much as possible, because Andrew’s ability to relax and focus on his work was strongly related to their own ability to stay relaxed. Participants were also informed that the whole experiment would last for about 45 minutes. The version of the level employed in the evaluation was modified in such a way that during the game session no game over screen was shown, and the level score was never displayed. Furthermore, we did not mention the score and its associated concept (i.e., percentage of writing task completed by Andrew) to participants.

After the introductory briefing, all participants consented to participate to the experiment. They were seated, the skin of their forehead and cheeks was cleaned using a pad of cotton wool and alcohol, and all the physiological sensors were applied.

Before each condition, participants carried out a baseline session, during which they were asked to rest for a minute, so that physiological parameters could revert to a rest state. Also, the baseline recording data was employed in the multi–sensor condition to normalize physiological data. After each baseline recording, participants carried out the game session in one of the three experimental conditions for 3 minutes. To prevent learning effects, we counterbalanced the order of the three conditions in such a way that every six participants, all the six possible combinations of the three conditions were tested, and that each condition was tested the same number of times as first, second, and third condition during the study. During each game session, physiological data were recorded from all the sensors. Then, after each game session, they filled the questionnaire for assessing quality of biofeedback and difficulty of relaxation for the game session just tried.

After trying all three game sessions, participants were asked for possible comments about each. Finally, all the physiological sensors were removed, and participants were debriefed and thanked for their participation.

8.4.4 Hypothesis

For perceived quality of biofeedback, we hypothesized that the two conditions (EDA–only, multi–sensor) that rely on a true affective computing approach to biofeedback would receive better ratings than the sham feedback condition (placebo). In addition, since the computation of user’s stress performed in the multi–sensor condition takes into account more physiological signals, we expected it to improve quality with respect to EDA–only. The inclusion of the difficulty of relaxation training measure was more exploratory in nature. On one hand, higher quality of biofeedback is likely to facilitate users in learning relaxation training. On the other hand, biofeedback–based relaxation training is not learned in a few minutes and our participants had no experience about it, so the differences in difficulty of relaxation training among conditions were likely to be small.

8.5 Results

We carried out a Shapiro–Wilk normality test on the questionnaire data, which indicated that they follow a Gaussian distribution. Then, we performed a repeated measures multivariate analysis of variance (MANOVA). When multiple dependent variables are measured as in our case, a single MANOVA is an appropriate statistical test, because it helps control for the inflation of Type I error that characterizes the use of multiple ANOVAs (Coolican, 2009). In our case, a Pearson correlation analysis among questionnaire data shows moderate correlations between perceived quality of biofeedback scores and difficulty of relaxation training scores (the higher the perceived quality of biofeedback, the smaller the difficulty of relaxation) for the multi-sensor condition ($\rho = -0.41$, $p < 0.05$) and the EDA-only condition ($\rho = -0.43$, $p < 0.05$). MANOVA revealed a main effect of stress detection technique (Wilks' $\lambda = 0.86$, $F(2, 68) = 2.58$, $p < 0.05$, $\eta_p^2 = 0.07$). We then proceeded with univariate tests with Greenhouse–Geisser correction and Bonferroni correction of significance level (α set at 0.025), which showed significant differences in perceived quality of biofeedback ($p < 0.01$, $\eta_p^2 = 0.14$). Pairwise comparisons (with Bonferroni correction) among the feedback scores in the three conditions revealed a significant difference between the EDA-only ($M = 5.00$, $SD = 1.09$) and the placebo ($M = 4.30$, $SD = 1.39$) conditions ($p < 0.01$), while no significant difference could be observed between the EDA-only and multi-sensor ($M = 4.55$, $SD = 1.38$) conditions, or the multi-sensor and the placebo conditions.

Figure 8.4 shows the mean values of the feedback score.

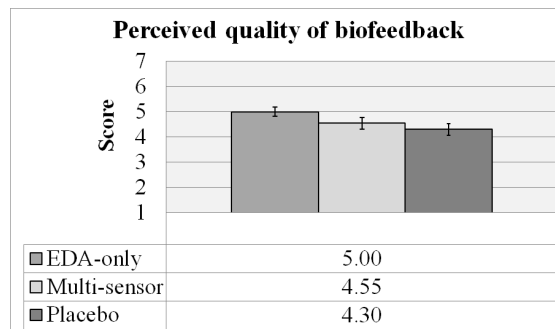


Figure 8.4: Mean values of the perceived quality of the biofeedback score. Error bars indicate standard error of the mean.

The differences (Figure 8.5) in the mean ratings for the difficulty of relaxation training with EDA-only ($M = 3.14$, $SD = 1.35$), multi-sensor ($M = 3.22$, $SD = 1.34$) and placebo ($M = 3.32$, $SD = 1.25$) were small and not statistically significant.

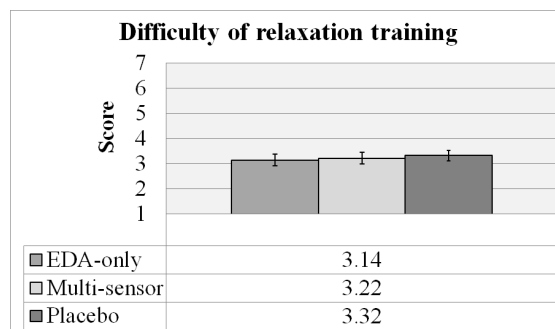


Figure 8.5: Mean values of the difficulty of relaxation training score. Error bars indicate standard error of the mean.

8.6 Discussion of the Results

For perceived quality of biofeedback, the results partially confirmed our hypothesis.

As expected, the EDA-only algorithm performed significantly better than the placebo. Contrary to our expectations, the multi-sensor algorithm did not obtain better ratings with respect to EDA-only. This can be possibly explained by two considerations.

First, with regards to arousal, EDA may be sufficient for the detection of cognitively-related arousal increments (Boucsein, 2006), as discussed in Section 1.3.1, and HR might not be able to detect the more subtle arousal level changes induced by the biofeedback application, because it is an indicator for the higher arousal range (see Section 1.3.3).

Second, the real-time detection of valence for high-arousal affective states through the use of facial EMG may not have contributed to improve stress detection in the case of the proposed game. In particular, we can consider that the employed acoustic (ringing phone) and visual stimuli (Andrew's body animations and facial expressions) are unlikely to generate positive affect in participants in any of the three conditions, because they are not pleasant and they also distract them from the goal of reaching the maximal level of relaxation.

However, the very surprising result was that the multi-sensor algorithm lacked significant differences also from the placebo condition. The EDA signal, compared to EMG and HR, seems thus to have been useful in producing an arousal index which was more directly related to the stress level of participants playing the game. In general, the fuzzy rules by Mandryk and Atkins (2007), which were defined to detect the level of five different emotions and tested during play of competitive sports video games, might not be well suited to the context of the present study, in which we focus on detecting stress during a relaxation training activity. The different way in which the EDA signal is analyzed by the two algorithms might also have played a role. The results seem to suggest that inferring increases or decreases of the arousal level from the slope of the EDA signal could lead to a higher accuracy in arousal detection with respect to placebo, unlike using a unique baseline value calculated before the 3-min gaming session as a reference, which was the case in the multi-sensor algorithm. We can hypothesize that the use of a unique baseline value is not always able to correctly capture the trend of the player's arousal level. Let us consider, for example, a significant increase of EDA (i.e., greater than $0.05 \mu S$) after a long period during which the signal continued to decrease starting from the baseline (e.g., this happens when the player continuously increases his/her relaxation level since the start of the game). In the multi-sensor algorithm, the EDA increase cannot be detected until the signal value has increased enough to exceed the baseline, which may require time. The same holds true in the case of a player that starts relaxing after increasing his/her arousal level for a long time. With the EDA-only algorithm, these EDA increases and decreases can be detected immediately.

It is interesting to note that if the experiment had focused only on comparing the two non-placebo conditions, it would have concluded that there are no significant differences between the EDA-only and the multi-sensor algorithms, possibly giving the impression that using one stress detection solution or the other would produce similar results in the biofeedback game. The introduction of a placebo condition in the experimental method put instead results under a different light. It is important to highlight that the placebo condition employed in our study is different from control conditions used in affective computing and HCI literature. In control conditions, it is clear to participants that the presented feedback does not take into account their physiological signals (e.g., Rezazadeh et al., 2012). In other cases, the control condition is designed as a baseline measure, with the absence of stressful stimuli to make participants as much relaxed as possible (e.g., Nasoz et al., 2010). In our case, in the placebo condition participants were led to believe that their physiological signals were measured and used by the application. Moreover, the stressful stimuli were the same as the other conditions.

In our experiment, the EDA-only condition was able to display stress level to participants with significantly better accuracy than the placebo condition, while the multi-sensor condition was not able to do the same. Therefore, faced with the choice of which solution to employ in the biofeedback game, the results indicate that the EDA-only solution is really superior to sham feedback that mainly appeals to user's suggestion, while the (more costly and complex) multi-sensor condition was not able to show this superiority. Therefore, only the EDA-only solution can be justified in the considered case. More generally, this highlights the importance of using more thorough methodologies in future affective computing studies,

including placebo conditions to determine if the (often complex and costly) traditional methodologies for the evaluation of ACSs are actually playing a significant role in determining the effectiveness of the feedback. In affective computing studies, methodologies employing instruments and measures like confusion matrices (e.g., Healey and Picard, 2005; Zhai and Barreto, 2006) or mean successful recognition (e.g., Liu et al., 2009; Wu et al., 2010) allow researchers to assign an absolute value to the accuracy of stress detection. However, these procedures do not account for the placebo effect that can take place when the proposed emotion detection techniques are applied, for example, to biofeedback–based relaxation training, including placebo conditions in the experiment can thus complement the currently used methods to determine the accuracy of an ACS as well as its effectiveness in specific applications. Indeed, if an affective computing application could not exhibit a better accuracy than a placebo condition, it becomes difficult to justify the usage of complex (and often costly and inconvenient) affect detection techniques in the application. Furthermore, if a placebo condition had been considered, studies in the affective computing field which compared two or more systems might have gained more insights, especially with reference to the practical usefulness of the considered systems.

The results obtained with difficulty of relaxation training display a trend similar to the one obtained for quality of feedback, but the differences among conditions are in this case very small and not statistically significant. This indicates that, as suspected, difficulty of relaxation training is not likely to significantly benefit from the higher quality of biofeedback in a few minutes of first use of biofeedback–based relaxation training. Since we kept a recording of the physiological signals during the three conditions, we carried out an exploratory comparison that confirms the users' subjective indication. We analyzed the mean values of the tonic component of the EDA, extracted using the Ledalab software (Golz and Kaernbach, 2013), as well as the mean values of EMG and HR, recorded during the three game sessions. For all conditions, baseline values recorded before each condition (see Section 8.4.3) were subtracted from the physiological data recorded during the 3–min play session. Data was not normally distributed; therefore, we performed four separate nonparametric analyses of variance using Friedman tests. We did not obtain statistically significant differences, which was largely expected after just one session with each version of the game.

Subjective difficulty as well as physiological effectiveness of training need to be studied in the context of a realistic relaxation training course in which participants carry out multiple sessions over several weeks or months, e.g., as in (Buckelew et al., 1998; Foster, 2004; Jensen et al., 2009). However, the correlation between perceived quality of feedback and difficulty of relaxation training reported in Section 8.5 is promising.

Conclusions

This thesis highlighted the importance of physiology for different research topics other than affective computing, illustrating how physiological signals can be successfully employed to evaluate the ergonomics of novel interaction devices and techniques like LPS interaction (Chapter 3) to better understand its effects on users' musculoskeletal system. Physiology was also useful for conducting the evaluation of a violent video game that gave more prominence to the effects on UX with respect to classic studies on video game violence (Chapter 4).

This thesis also showed that physiology can be used to evaluate the effectiveness of interactive visualizations in breathing training apps, by measuring the effects on participants' breathing pattern (Chapter 6). To be able to influence users' emotions on the basis of their bodily reactions, we then developed an ACS for automatic stress detection that builds upon the current state of the art in affective computing (Chapter 7). The proposed ACS, called ACME, can read in real-time users' physiological signals, and determine their level of stress. ACME was tested on a game designed to support relaxation training (Chapter 8) that employs realistic 3D graphics to provide players with embodied feedback.

Affective computing, however, needs to further progress: as reported by Fairclough (2009), there are still issues that have a critical impact on the development and evaluation of affective computing systems (ACSs), and many ACSs proposed in the literature (Chapter 2) show less than optimal emotion detection accuracy. A possible solution to improve the detection of stress and anxiety could be the use of signals that are successfully employed in other research fields, such as eye-blink startle response (EBSR). Chapter 5 of this thesis shows how EBSR could be used in natural ways to extend the set of physiological measures currently employed in affective computing studies. Finally, this thesis proposed an improvement over the traditional comparison methodology employed in ACSs evaluations in affective computing studies (Chapter 8). For the evaluation of the proposed game, we added a placebo condition in which user's stress level, unbeknown to him/her, is determined pseudo-randomly instead of taking into account his/her physiological sensor readings. The results of the study of our thesis show that without the placebo condition the effects of biofeedback based on ACME and on a second algorithm would have instead appeared similar.

In the following, we discuss in more detail the results of our work and outline future research directions for each one of the topics we have explored.

C.1 Physiology-Based Evaluation of Interaction Devices

We started our research by exploiting physiological measurements to deal with the evaluation of the ergonomics of a novel interaction device based on laser pointer-style (LPS) interaction. The literature usually focuses on the performance afforded by these novel interaction techniques and devices, using subjective ratings to assess task difficulty and, more rarely, perceived fatigue. The goal of the study described in Chapter 3 is to better understand the effects of LPS interaction on users' musculoskeletal system through physiological measurements.

The results indicate that the traditional mouse and keyboard (M&K) setup granted a better performance than the LPS setup based on the Wii Remote and the Nunchuck (W&N) in object arrangement, and that M&K was also the interaction technique preferred by participants. This is consistent with results reported in the literature by studies that compared the mouse to LPS devices (e.g., MacKenzie et al., 2001; Olsen Jr and Nielsen, 2001; Myers et al., 2002).

The electromyography (EMG) data produced interesting results, which provide new insights about LPS interaction. The EMG analysis we performed, seen in the light of findings from medical and ergonomics literature (e.g., Aarås, 1994; Rozmaryn, 2005; Sommerich et al., 2006; Scheumann, 2007), suggest that, for left and right extensor digitorum communis and for biceps brachii, the risk of *work-related upper extremity disorders* (WUEDs) and more generally of musculoskeletal disorders (MSDs) could be greater with W&N than M&K. Moreover, results from the analysis of superior trapezius reported a lower muscle activity with

W&N than M&K, but the standing posture caused a greater amount of superior trapezius activity with W&N than M&K. This result is particularly interesting since the standing posture should be the most advantageous one for LPS interaction with large displays because of the mobility afforded.

Our results might potentially concern any handheld LPS device, since these devices are likely to require movements and postures similar to those observed in our study (for example, continuously pointing at the screen). However, more advanced LPS interaction techniques assist users in selection and manipulation tasks in various ways: e.g., *InterSelect* (De Haan et al., 2005) estimates which object the user wants by employing selection volumes; *Bubble Cursor* (Grossman and Balakrishnan, 2006) resizes and reshapes dynamically the circular cursor so that it contains only one object; *SQUAD* (Kopper et al., 2011) uses a spherical cursor for a first selection step, allowing users to further refine the selection by discrete steps which discard unwanted objects. Advanced LPS interaction techniques could thus elicit a lower muscle activity and afford a better performance. Therefore, it is important to assess in depth the effects on the musculoskeletal system for each of these more advanced techniques.

Finally, we plan to further exploit physiological measurements in future studies by considering, for example, the activity of a greater number of muscles, to obtain more detailed information about the effects of novel interaction techniques on the musculoskeletal system.

C.2 Physiology–Based Evaluation of Video Game User Experience

Chapter 4 described an user evaluation that, unlike the one described in Chapter 3, focused on employing physiological measurements to infer users' emotional state and evaluate the user experience (UX) afforded by a violent mobile video game, on the basis of the literature reviewed in Chapter 1.

The results obtained from the analysis of participants' physiological data suggest that previous literature findings about effects (in terms of affect during play and subsequent desensitization to violence) of performing aggressive actions against virtual humans and anthropomorphic beings in video games may not necessarily generalize to aggressive actions towards non-human animal species. The act of killing (virtual) insects in the violent version of the Whac-a-Mole game seemed to be no more desensitizing than the non-violent one. Moreover, it appeared to be more entertaining, considering the higher left and right zygomaticus major activity during game play, which also indicates that replacing the violent element of Whac-a-Mole games could not be a straightforward game design task. It is worth remembering that, by using two versions of the same game that differed only in the action performed by the player (violent and non-violent), we could achieve a level of control on possibly confounding variables which can only be found in very few other papers in the literature (Bluemke et al., 2010; Hartmann et al., 2010).

Widespread social tolerance of aggressive acts against some animal species and difficulty to empathize with those species could be possible factors that explain the obtained results. It must be noted that, exactly like insects, moles (the animals used in the classic Whac-A-Mole arcade game) are also a socially tolerated target of unnecessary human violence (a simple Web search produces thousands of sites devoted to teach mole killing techniques or promote specific products and services to the purpose), despite appeals for a change in attitude toward moles by animal rights associations (Baird, 2009). A reason for the success of Whac-A-Mole games might thus be that they allow players to engage in violent acts against species that do not elicit the negative physiological responses (and affect) that the literature has instead highlighted for acts of killing or wounding virtual human beings. Notwithstanding the limitations of the exploratory study discussed in this chapter (see Section 4.6.4), the presented results suggest that further research in this field may lead to interesting findings. An increased number of participants in future studies may allow us to obtain more solid information about the desensitization effects of the considered violent video game genre, and to better compare our results with the existing literature. Furthermore, a more balanced gender distribution will allow us to explore differences in the effects on physiology of violent video game play in males and females, with regards to both UX and desensitization to violence.

Finally, we plan to evaluate the effects on player's physiology of different devices which should provide a greater sensory stimulation, e.g., by comparing large LCD monitors and surround sound vs. a mobile device to explore the possible role of the employed devices on the effects of violent video games.

C.3 Novel Approaches for Stress and Anxiety Detection

To improve the current body of knowledge about the use of physiological signals for the detection of stress- and anxiety-related affective states (Chapter 4), Chapter 5 investigated if EBSR, a physiological signal that is commonly employed in the psychophysiology literature, could be employed in affective computing studies. In particular, we focused on assessing the possibility of using more natural startle stimuli and encourage further exploration with different realistic and recognizable acoustic startle stimuli in place of white noise, which is the standard stimulus employed in the literature; the observed results seems to suggest that this is the case.

The ultimate goal would be to extend in a natural way the set of physiological measures employed in affective computing with EBSR, which can be used to assess the intensity of users' negative emotions such as anxiety (e.g., Grillon et al., 1991; Grillon, 2002; Baas et al., 2004), as well as sense of presence in the virtual environment (VE) (e.g., Nichols et al., 2000; Parsons et al., 2009). As shown by some studies (e.g., Zhai and Barreto, 2006), the use of multiple physiological sensors may help to improve the accuracy of automatic stress detection.

Sounds like explosions can naturally present themselves in real-world events, and thus can be easily interpreted in the context of a realistic VE, because they have a meaningful relation with the visual cues presented by the virtual experience. The presentation of a natural acoustic stimulus, unlike white noise, can therefore become an integral part of the user experience, and may even increase the immersion of the user instead of potentially hindering it. It is important, however, to accurately select the right natural sound for the chosen VE, otherwise the stimuli, albeit natural, would be as intrusive and distracting as white noise: for example, consider the explosion sound in a relaxing virtual experience inside a forest. In searching for a suitable natural stimuli, one must balance: (i) the realism and familiarity of the sound (i.e., how easily it can be recognized and interpreted as an element of the VE), and (ii) its effectiveness in eliciting a startle response of adequate intensity.

For these reasons, in future studies, we will evaluate different natural noises to create a collection of realistic startle stimuli that can be easily associated to virtual objects and events: besides explosions and crashes, one can consider sounds such as human or animal shouts, car horns, telephone rings, and so on. A large number of tested stimuli can extend the number of applications in which EBSR could be naturally applied.

Finally, we will extend our study by evaluating not only the magnitude of the responses elicited by different startle stimuli, but also the possible effects of different virtual experiences, focusing in particular on comparisons between stressful and non-stressful VEs (e.g., Mühlberger et al., 2008; Parsons et al., 2012). The goal is to assess if natural startle stimuli, besides being able to elicit startle responses that are comparable to white noise as we have seen in this chapter, have a different discriminating power in analyzing the valence of emotions.

C.4 Physiology-Based Evaluation of a Mobile Breathing Training Application

After focusing on novel interaction devices (Chapter 3) and video game UX (Chapter 4), in Chapter 6 we employed physiology for the evaluation of different designs for a mobile breathing training application. Breathing exercises have positive effects on blood pressure and hypertension and can be an effective adjunct in the treatment of stress, anxiety, post-traumatic stress disorder, chronic pain and depression. Breathing training applications, therefore, could play a significant role in relaxation training.

Our study, which employed electrodermal activity (EDA), heart rate (HR), and respiration as physiological signals, has provided experimental support to the recent proposal of using mobile apps for breathing training. If we can assume that the way breathing instructions are displayed is carefully designed, using a mobile app can allow trainees to obtain objectively better results in performing deep and slow respiration and they can also find the app more effective from an instruction as well as a relaxation point of view. Of the three app designs we evaluated, the one adopting a wave-based visualization was able to produce the better objective and subjective results. To the best of our knowledge, our study is the first to focus on

the evaluation of mobile apps for breathing training and to shed light on the positive role that computer visualization can play in such applications.

Thanks to the availability of mobile devices anytime and anywhere, trainees can have a convenient access to effective and affordable breathing training apps that could be integrated more easily in their daily routine with respect to desktop applications or traditional approaches (following courses given by instructors and using audio CDs). As we have seen, a wave-based mobile visualization also allows to obtain better results with respect to audio-only instructions. This should be taken into account in interventions (e.g., for hypertension or stress-related disorders) which adopt phones as a m-health instrument for patient training.

To evaluate the long-term effects of different apps, the next phase in our research will concern the design and execution of a longitudinal study to gain broader insights about (i) how the applications are used by trainees in their daily life, (ii) the possible differences in facilitating the transition from app-supported respiration to independent application of the breathing techniques in everyday life, and (iii) by exploiting users' physiology, the possible differences in the levels of relaxation and wellbeing obtained, e.g., users' ability to keep their stress level under control.

Finally, future studies will also explore further the design space of visualizations, by considering improvements to the two types of visualizations evaluated in the current study as well as extending our attention to other visualizations we described in Section 6.2. In particular, we are interested to assess if animations that resort to human body visualization might further improve the results obtained in the current study.

C.5 Physiology for Building and Evaluating a Biofeedback-Based Relaxation Training Application

Unlike the training application proposed and evaluated in Chapter 6, Chapter 8 introduced an application that combines ideas from affective computing, biofeedback and serious games to provide richer visual and auditory feedback, presenting users with real-life simulated scenarios in a 3D environment and using the behavior of a virtual character in the environment to represent user's physiological state. The study presented in Chapter 8 evaluated the accuracy of the stress detection afforded by ACME, the ACS proposed in Chapter 7, by comparing it with a second system based on an algorithm that only employs EDA to detect users' stress. Furthermore, unlike most research in the literature, the user evaluation presented in Chapter 8 compared the two stress detection systems with a sham biofeedback condition (placebo), in which users stress level, unbeknown to them, is determined pseudo-randomly instead of taking into account their current physiological sensor readings.

Results of the evaluation showed that only the EDA-only approach was able to produce a feedback that was perceived as significantly more accurate than the placebo condition. If only the two non-placebo conditions had been considered, their effectiveness would have instead appeared similar. Consideration of the placebo condition put results in another perspective, pointing out that only in one of the conditions the exploitation of physiological sensors contributed to significantly improve the quality of feedback. In general, this outcome highlights the importance of using more thorough methodologies in future affective computing studies, including placebo conditions.

The next phase in our research will concern the design and execution of a longitudinal study. While a different quality of feedback can be noticed by users in a short period of time as we have seen in the present study, a longitudinal study would allow us to assess further possible advantages of a better feedback, e.g., in terms of easier and faster learning of relaxation training and/or a better outcome of the relaxation training treatment. To better test ACME capabilities, we will design new levels for the biofeedback game described in Section 8.3, aiming at presenting users also with acoustic and visual stimuli that can elicit positive emotions.

This might allow ACME to gain an advantage over the EDA-only algorithm, because it could possibly be able to distinguish between emotional states characterized by negative high-arousal levels (which is typical of a stressful state), and positive high-arousal levels (a state of excitement unrelated to stress). The ability to distinguish between these two high-arousal states could allow the affective computing application to influence different aspects of the users' affective state: in the negative high arousal case, the training

activity must be able to reduce both arousal and the negativity of emotions; in the positive high arousal case, the training activity should focus on reducing the arousal, while maintaining positive emotional states, which could optimize health and well-being (Fredrickson, 2000).

Bibliography

- 3CUBE (2012). ColorBREATH. <https://itunes.apple.com/us/app/colorbreath/id522637619?mt=8> (last accessed on 11 June 2013).
- Aarås, A. (1994). Relationship between trapezius load and the incidence of musculoskeletal illness in the neck and shoulder. *International Journal of Industrial Ergonomics*, 14(4):341–348.
- Accot, J. and Zhai, S. (1997). Beyond Fitts' law: models for trajectory-based HCI tasks. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 1997)*, pages 295–302. ACM Press.
- Adachi, P. J. C. and Willoughby, T. (2011). The effect of violent video games on aggression: is it more than just the violence? *Aggression and Violent Behavior*, 16(1):55–62.
- Akritas, M. G. and Brunner, E. (1997). A unified approach to rank tests for mixed models. *Journal of Statistical Planning and Inference*, 61(2):249–277.
- Akselrod, S., Gordon, D., Ubel, F. A., Shannon, D. C., Berger, A. C., and Cohen, R. J. (1981). Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control. *Science*, 213(4504):220–222.
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*. DSM-IV-TR. American Psychiatric Association, 4th edition.
- Anderson, C. A. and Carnagey, N. L. (2009). Causal effects of violent sports video games on aggression: is it competitiveness or violent content? *Journal of Experimental Social Psychology*, 45(4):731–739.
- Anderson, C. A., Carnagey, N. L., and Eubanks, J. (2003). Exposure to violent media: the effects of songs with violent lyrics on aggressive thoughts and feelings. *Journal of Personality and Social Psychology*, 84(5):960–971.
- Anderson, C. A., Carnagey, N. L., Flanagan, M., Benjamin Jr., A. J., Eubanks, J., and Valentine, J. C. (2004). Violent video games: specific effects of violent content on aggressive thoughts and behavior. *Advances in Experimental Social Psychology*, 36:199–249.
- Anderson, C. A., Deuser, W. E., and DeNeve, K. M. (1995). Hot temperatures, hostile affect, hostile cognition, and arousal: tests of a general model of affective aggression. *Personality and Social Psychology Bulletin*, 21(5):434.
- Anderson, C. A. and Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology*, 78(4):772–790.
- Anderson, C. A. and Ford, C. M. (1986). Affect of the game player: short term effects of highly and mildly aggressive video games. *Personality and Social Psychology Bulletin*, 12(4):390–402.
- Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., Rothstein, H. R., and Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: a meta-analytic review. *Psychological Bulletin*, 136(2):151–173.
- Andreassi, J. L. (2007). *Psychophysiology — Human Behavior & Physiological Response*. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 5th edition.
- Arns, M., De Ridder, S., Strehl, U., Breteler, M., and Coenen, T. (2009). Efficacy of neurofeedback treatment in ADHD: The effects on inattention, impulsivity and hyperactivity: A meta-analysis. *Clinical EEG & Neuroscience*, 40(3):180–189.

- Arriaga, P., Esteves, F., Carneiro, P., and Monteiro, M. B. (2006). Violent computer games and their effects on state hostility and physiological arousal. *Aggressive Behavior*, 32(4):358–371.
- Arriaga, P., Esteves, F., Carneiro, P., and Monteiro, M. B. (2008). Are the effects of unreal violent video games pronounced when playing with a virtual reality system? *Aggressive Behavior*, 34(5):521–538.
- Autodesk (2013). 3ds Max. <http://www.autodesk.it/products/autodesk-3ds-max/overview> (last accessed on 17 December 2013).
- Baas, J. M. P., Nugent, M., Lissek, S., Pine, D. S., and Grillon, C. (2004). Fear conditioning in virtual reality contexts: a new tool for the study of anxiety. *Biological Psychiatry*, 55(11):1056–1060.
- Baird, M. (2009). Wild things: moles. *Animal Sheltering*, March–April 2009:16.
- Baños, R. M., Guillen, V., Quero, S., García-Palacios, A., Alcañiz, M., and Botella, C. (2011). A virtual reality system for the treatment of stress-related disorders: a preliminary analysis of efficacy compared to a standard cognitive behavioral program. *International Journal of Human–Computer Studies*, 69(9):602–613.
- Barlett, C. P., Harris, R. J., and Bruey, C. (2008). The effect of the amount of blood in a violent video game on aggression, hostility, and arousal. *Journal of Experimental Social Psychology*, 44(3):539–546.
- Barlow, D. H. (1991). Disorders of emotions. *Psychological Inquiry*, 2(1):58–71.
- Baron, R. A., Byrne, D. R., and Branscombe, N. R. (2009). *Social Psychology*. Pearson, Boston, MA, USA.
- Bartholow, B. D., Bushman, B. J., and Sestir, M. A. (2006). Chronic violent video game exposure and desensitization to violence: behavioral and event-related brain potential data. *Journal of Experimental Social Psychology*, 42(4):532–539.
- Bassotti, G. and Whitehead, W. E. (1997). Biofeedback, relaxation training, and cognitive behaviour modification as treatments for lower functional gastrointestinal disorders. *Quarterly Journal of Medicine*, 90(8):545–550.
- Becker, C., Zayfert, C., and Anderson, E. R. (2004). A survey of psychologists' attitudes towards and utilization of exposure therapy for PTSD. *Behaviour Research and Therapy*, 42:277–292.
- Benchetrit, G. (2000). Breathing pattern in humans: diversity and individuality. *Respiration Physiology*, 122:123–129.
- Benedek, M. and Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47(4):647–658.
- Benson, H. and Klipper, M. Z. (1975). *The relaxation response*. HarperCollins.
- Bérard, F., Ip, J., Benovoy, M., El-Shimy, D., Blum, J. R., and Cooperstock, J. (2009). Did “Minority Report” get it wrong? Superiority of the mouse over 3D input devices in a 3D placement task. In *Proc. 12th IFIP TC13 Conference on Human–Computer Interaction (INTERACT 2009)*, pages 400–414. Springer.
- Bersak, D., McDarby, G., Augenblick, N., McDarby, P., McDonnell, D., McDonald, B., and Karkun, R. (2001). Intelligent biofeedback using an immersive competitive environment. In *Online proc. Designing Ubiquitous Computing Games Workshop, UbiComp 2001*. Springer.
- Bluemke, M., Friedrich, M., and Zumbach, J. (2010). The influence of violent and nonviolent computer games on implicit measures of aggressiveness. *Aggressive Behavior*, 36(1):1–13.
- Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., and van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, 42(1):1–15.

- Bolt, R. A. (1980). "Put-that-there": voice and gesture at the graphics interface. *ACM Siggraph Computer Graphics*, 14(3):262–270.
- Borkovec, T. D. (2002). Life in the future versus life in the present. *Clinical Psychology: Science and Practice*, 9(1):76–80.
- Borkovec, T. D., Shadick, R. N., and Hopkins, M. (1991). The Nature of Normal and Pathological Worry. In Rapee, R. M. and Barlow, D. H., editors, *Chronic anxiety: Generalized anxiety disorder and mixed anxiety-depression*, pages 29–51. The Guilford Press, New York, NY, USA.
- Bouchard, S., Bernier, F., Boivin, É., Morin, B., and Robillard, G. (2012). Using biofeedback while immersed in a stressful videogame increases the effectiveness of stress management skills in soldiers. *PLoS ONE*, 7(4).
- Boucsein, W. (2006). *Electrodermal Activity*. Springer, New York, NY, USA, 2nd edition.
- Bowman, D. A. and Hodges, L. F. (1997). An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. *Proc. of the 1997 Symposium on Interactive 3D Graphics (I3D 1997)*, pages 35–ff.
- Bowman, D. A., Koller, D., and Hodges, L. F. (1997). Travel in immersive virtual environments: an evaluation of viewpoint motion control techniques. *Proc. Virtual Reality Annual International Symposium (VRAIS 1997)*, pages 45–52, 215.
- Boyle, E., Connolly, T. M., and Hainey, T. (2011). The role of psychology in understanding the impact of computer games. *Entertainment Computing*, 2(2):69–74.
- Bradley, M. M., Cuthbert, B. N., and Lang, P. J. (1990). Startle reflex modification: emotion or attention? *Psychophysiology*, 27(5):513–522.
- Bradley, M. M. and Lang, P. J. (2007). Emotion and Motivation. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G., editors, *Handbook of Psychophysiology*, pages 581–607. Cambridge University Press, New York, NY, USA.
- Breathing Zone (2011). Breathing Zone. <http://www.breathing-zone.com/> (last accessed on 11 June 2013).
- Brouwer, A.-M., Neerinx, M. A., Kallen, V., van der Leer, L., and ten Brinke, M. (2011). EEG alpha asymmetry, heart rate variability and cortisol in response to virtual reality induced stress. *Journal of CyberTherapy & Rehabilitation*, 4(1):21–34.
- Brown, R. P. and Gerbarg, P. L. (2005). Sudarshan Kriya Yogic breathing in the treatment of stress, anxiety, and depression. Part II — clinical applications and guidelines. *Journal of alternative and complementary medicine (New York, N.Y.)*, 11(4):711–717.
- Brunner, E., Munzel, U., and Puri, M. L. (1999). Rank–score tests in factorial designs with repeated measures. *Journal of Multivariate Analysis*, 70(2).
- Buckelew, S. P., Conway, R., Parker, J., Deuser, W. E., Read, J., Witty, T. E., Hewett, J. E., Minor, M., Johnson, J. C., and Van Male, L. (1998). Biofeedback/relaxation training and exercise interventions for fibromyalgia: a prospective trial. *Arthritis & Rheumatism*, 11(3):196–209.
- Busch, V., Magerl, W., Kern, U., Haas, J., Hajak, G., and Eichhammer, P. (2012). The effect of deep and slow breathing on pain perception, autonomic activity, and mood processing—an experimental study. *Pain medicine (Malden, Mass.)*, 13(2):215–228.
- Bushman, B. J. and Anderson, C. A. (2002). Violent video games and hostile expectations: a test of the general aggression model. *Personality and Social Psychology Bulletin*, 28(12):1679.
- Buss, A. H. and Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, 63(3):452–459.

- Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., and Ito, T. A. (2000). The psychophysiology of emotion. In *Handbook of Emotions*, pages 173–191. The Guilford Press, New York, NY, USA.
- Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G., editors (2007). *Handbook of Psychophysiology*. Cambridge University Press, New York, NY, USA, 3rd edition.
- Campbell, B. A., O'Brien, K. R., Byrne, M. D., and Bachman, B. J. (2008). Fitts' law predictions with an alternative pointing device (wiimote®). In *Proc. of the Human Factors and Ergonomics Society Annual Meeting (HFES 2008)*, volume 52, pages 1321–1325.
- Cannon, W. B. (1927). The James–Lange theory of emotions: a critical examination and an alternative theory. *American Journal of Psychiatry*, 39(1/4):106–124.
- Caprara, G. V., Cinanni, V., D'Imperio, G., Passerini, S., Renzi, P., and Travaglia, G. (1985). Indicators of impulsive aggression: present status of research on irritability and emotional susceptibility scales. *Personality and Individual Differences*, 6(6):665–674.
- Carnagey, N. L., Anderson, C. A., and Bushman, B. J. (2007). The effect of video game violence on physiological desensitization to real–life violence. *Journal of Experimental Social Psychology*, 43(3):489–496.
- Carr, A. (2006). *The Handbook of Child and Adolescent Clinical Psychology: A Contextual Approach*. Routledge, London, UK, 2nd edition.
- Casiez, G., Vogel, D., Balakrishnan, R., and Cockburn, A. (2008). The impact of control–display gain on user performance in pointing tasks. *Human–Computer Interaction*, 23(3):215–250.
- Cavens, D., Vogt, F., Fels, S., and Meitner, M. (2002). Interacting with the big screen: pointers to ponder. In *Proc. Extended Abstracts on Human Factors in Computing Systems (CHI 2002)*, pages 678–679. ACM Press.
- Chandler, C., Bodenhamer Davis, E., Holden, J. M., Evenson, T., and Bratton, S. (2001). Enhancing personal wellness in counselor trainees using biofeedback: an exploratory study. *Applied psychophysiology and biofeedback*, 26(1):1–7.
- Chanel, G., Rebetez, C., Bétrancourt, M., and Pun, T. (2011). Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Transactions on Systems, Man, and Cybernetics — Part A: Systems and Humans*, 41(6):1052–1063.
- Cheng, K. and Pulo, K. (2003). Direct interaction with large–scale display systems using infrared laser tracking devices. In *Proc. of the Asia–Pacific Symposium on Information Visualisation (APVis 2003) — Volume 24*, pages 67–74. Australian Computer Society, Inc.
- Chittaro, L. (2001). Information visualization and its application to medicine. *Artificial Intelligence in Medicine*, 22(2):81–88.
- Chittaro, L. (2006). Visualizing information on mobile devices. *IEEE Computer*, 39(3):40–45.
- Chittaro, L., Ranon, R., and Ieronutti, L. (2009). 3D object arrangement for novice users: the effectiveness of combining a first–person and a map view. In *Proc. 16th ACM Symposium on Virtual Reality Software and Technology (VRST 2009)*, pages 171–178. ACM Press.
- Chittaro, L. and Sioni, R. (2012a). An electromyographic study of a laser pointer–style device vs. mouse and keyboard in an object arrangement task on a large screen. *International Journal of Human–Computer Studies*, 70(3):234–255.
- Chittaro, L. and Sioni, R. (2012b). Killing non–human animals in video games: a study on user experience and desensitization to violence aspects. *PsychNology Journal*, 10(3):215–243.

- Chittaro, L. and Sioni, R. (2013). Exploring eye–blink startle response as a physiological measure for affective computing. In *Proc. 5th International Conference on Affective Computing and Intelligent Interaction (ACII 2013)*, pages 227–232. IEEE Computer Society.
- Chittaro, L. and Sioni, R. (2014). Affective computing vs. affective placebo: Study of a biofeedback–controlled game for relaxation training. *International Journal of Human–Computer Studies*. DOI: 10.1016/j.ijhcs.2014.01.007.
- Clarkson, H. M. (2000). *Musculoskeletal Assessment: Joint Range of Motion and Manual Muscle Strength*. Lippincott Williams & Wilkins, Baltimore, MD, USA.
- Cohen, B. H. (2001). *Explaining Psychological Statistics*. Wiley, Hoboken, NJ, USA.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum, Hillsdale, NJ, USA.
- Conrad, A., Müller, A., Doberenz, S., Kim, S., Meuret, A. E., Wollburg, E., and Roth, W. T. (2007). Psychophysiological effects of breathing instructions for stress management. *Applied Psychophysiology and Biofeedback*, 32(2):89–98.
- Coolican, H. (2009). *Research Methods and Statistics in Psychology*. Hodder Arnold.
- Cornwell, B. R., Johnson, L., Berardi, L., and Grillon, C. (2006). Anticipation of public speaking in virtual reality reveals a relationship between trait social anxiety and startle reactivity. *Biological Psychiatry*, 59(7):664–666.
- Cox, E. (1992). Fuzzy fundamentals. *IEEE Spectrum*, 29(10):58–61.
- Critchley, H. D., Melmed, R. N., Featherstone, E., Mathias, C. J., and Dolan, R. J. (2001). Brain activity during biofeedback relaxation: a functional neuroimaging investigation. *Brain*, 124(Pt 5):1003–1012.
- Critchley, H. D., Melmed, R. N., Featherstone, E., Mathias, C. J., and Dolan, R. J. (2002). Volitional control of autonomic arousal: a functional magnetic resonance study. *Neuroimage*, 16(4):909–919.
- Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., and Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, 7(2):189–195.
- Cruz Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., and Hart, J. C. (1992). The CAVE: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–42.
- Csikszentmihályi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper & Row, New York, NY, USA.
- Davis, J. and Chen, X. (2002). LumiPoint: multi–user laser–based interaction on large tiled displays. *Displays*, 23(5):205–211.
- Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1):113–126.
- Dawson, M. E., Schell, A. M., and Filion, D. L. (2007). The Electrodermal System. In *Handbook of Psychophysiology*, pages 200–223. Cambridge University Press, New York, NY, USA.
- De Haan, G., Koutek, M., and Post, F. H. (2005). IntenSelect: using dynamic object rating for assisting 3D object selection. In *Proc. 11th Eurographics Conference on Virtual Environments (EGVE 2005)*, pages 201–209. Eurographics Association.
- De Silva, P. and Rachman, S. (1981). Is exposure a necessary condition for fear–reduction? *Behaviour Research and Therapy*, 19(3):227–232.
- Dekker, A. and Champion, E. (2007). Please biofeed the zombies: enhancing the gameplay and display of a horror game using biofeedback. In *DiGRA*, pages 1–9. Citeseer.

- Delgado, L. C., Guerra, P., Perakakis, P., Mata, J. L., Nieves Pérez, M., and Vila, J. (2009). Psychophysiological correlates of chronic worry: cued versus non-cued fear reaction. *International Journal of Psychophysiology*, 74(3):280–287.
- Delgado, L. C., Guerra, P., Perakakis, P., Vera, M. N., del Paso, G. R., and Vila, J. (2010). Treating chronic worry: psychological and physiological effects of a training programme based on mindfulness. *Behaviour Research and Therapy*, 48(9):873–882.
- Designit (2009). De Stress. <http://healthmate.eu/destress> (last accessed on 11 June 2013).
- Douglas, S. A., Kirkpatrick, A. E., and MacKenzie, I. S. (1999). Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. In *Proc. ACM SIGCHI Conference on Human factors in Computing Systems (CHI 1999)*, pages 215–222. ACM press.
- Drachen, A., Nacke, L. E., Yannakakis, G. N., and Pedersen, A. L. (2010). Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In *Proc. 5th ACM SIGGRAPH Symposium on Video Games (Sandbox 2010)*, pages 49–54. ACM Press.
- ECNA LAB (2011). ECNA–Breath. <https://itunes.apple.com/us/app/ecna-breath/id386514251?mt=8> (last accessed on 11 June 2013).
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In Cole, J., editor, *Nebraska Symposium on Motivation, 1971*, pages 207–282. University of Nebraska Press.
- Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2):124–129.
- Ekman, P., Levenson, R. W., and Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221(4616):1208–1210.
- Elliott, W. J., Izzo, J. L., White, W. B., Rosing, D. R., Snyder, C. S., Alter, A., Gavish, B., and Black, H. R. (2004). Graded blood pressure reduction in hypertensive outpatients associated with use of a device to assist with slow breathing. *The Journal of Clinical Hypertension*, 6(10):553–559.
- Elmqvist, N. and Fekete, J.-D. (2008). Semantic pointing for object picking in complex 3D environments. In *Proc. Graphics Interface 2008 (GI 2008)*. Canadian Information Processing Society.
- Enck, P. (1993). Biofeedback training in disordered defecation. *Digestive diseases and sciences*, 38(11):1953–1960.
- Engelhardt, C. R., Bartholow, B. D., Kerr, G. T., and Bushman, B. J. (2011). This is your brain on violent video games: neural desensitization to violence predicts increased aggression following violent video game exposure. *Journal of Experimental Social Psychology*, 47(5):1033–1036.
- Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interacting with Computers*, 21(1–2):133–145.
- Fairclough, S. H. and Goodwin, L. (2007). The effect of psychological stress and relaxation on interoceptive accuracy: implications for symptom perception. *Journal of Psychosomatic Research*, 62(3):289–295.
- Feldman, G., Greeson, J., and Senville, J. (2010). Differential effects of mindful breathing, progressive muscle relaxation, and loving-kindness meditation on decentering and negative reactions to repetitive thoughts. *Behaviour Research and Therapy*, 48(10):1002–1011.
- Fentress, D. W., Masek, B. J., Mehegan, J. E., and Benson, H. (1986). Biofeedback and relaxation-response training in the treatment of pediatric migraine. *Developmental Medicine & Child Neurology*, 28:139–146.
- Ferguson, C. J. (2007). The good, the bad and the ugly: a meta-analytic review of positive and negative effects of violent video games. *Psychiatric Quarterly*, 78(4):309–316.

- Fleureau, J., Guillotel, P., and Huynh-Thu, Q. (2012). Physiological-based affect event detector for entertainment video applications. *IEEE Transactions on Affective Computing*, 3(3):1–8.
- Foster, P. S. (2004). Use of the Calmset 3 biofeedback/relaxation system in the assessment and treatment of chronic nocturnal bruxism. *Applied psychophysiology and biofeedback*, 29(2):141–147.
- Fredrickson, B. L. (2000). Cultivating positive emotions to optimize health and well-being. *Prevention & Treatment*, 3:1–25.
- Fridlund, A. J. and Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, 23(5):567–589.
- Friedman, B. H. (2010). Feelings and the body: the Jamesian perspective on autonomic specificity of emotion. *Biological psychology*, 84(3):3–13.
- Friedman, D., Suji, K., and Slater, M. (2007). SuperDreamCity: an immersive virtual reality experience that responds to electrodermal activity. In *Affective Computing and Intelligent Interaction*, volume 4738 of *Lecture Notes in Computer Science*, pages 570–581.
- Fuller, B. F. (1992). The effects of stress–anxiety and coping styles on heart rate variability. *International Journal of Psychophysiology*, 12(1):81–86.
- Globisch, J., Hamm, A. O., Esteves, F., and Öhman, A. (1999). Fear appears fast: temporal course of startle reflex potentiation in animal fearful subjects. *Psychophysiology*, 36(1):66–75.
- Gl.tter (2007). WiiYourself!. <http://wiiyourself.gl.tter.org/> (last accessed on 19 June 2013).
- Golz, J. and Kaernbach, C. (2013). Ledalab. <http://www.ledalab.de> (last accessed on 25 october 2013).
- Gorini, A., Pallavicini, F., Algeri, D., Repetto, C., Gaggioli, A., and Riva, G. (2010). Virtual reality in the treatment of generalized anxiety disorders. *Studies in Health Technology and Informatics*, 154:1–5.
- Gorini, A. and Riva, G. (2008). The potential of Virtual Reality as anxiety management tool: a randomized controlled study in a sample of patients affected by Generalized Anxiety Disorder. *Trials*, 9(1):25–33.
- Graziano, M. S. A. and Cooke, D. F. (2006). Parieto–frontal interactions, personal space, and defensive behavior. *Neuropsychologia*, 44(6):845–859.
- Grillon, C. (2002). Startle reactivity and anxiety disorders: aversive conditioning, context, and neurobiology. *Biological Psychiatry*, 52(10):958–975.
- Grillon, C. and Ameli, R. (1998). Effects of threat of shock, shock electrode placement and darkness on startle. *International Journal of Psychophysiology*, 28(3):223–231.
- Grillon, C., Ameli, R., Woods, S. W., Merikangas, K., and Davis, M. (1991). Fear–potentiated startle in humans: effects of anticipatory anxiety on the acoustic blink reflex. *Psychophysiology*, 28(5):588–595.
- Grillon, C., Pellowski, M., Merikangas, K. R., and Davis, M. (1997). Darkness facilitates the acoustic startle reflex in humans. *Biological Psychiatry*, 42(6):453–460.
- Grossman, E., Grossman, A., Schein, M. H., Zimlichman, R., and Gavish, B. (2001). Breathing–control lowers blood pressure. *Journal of human hypertension*, 15(4):263–269.
- Grossman, P. and Taylor, E. W. (2007). Toward understanding respiratory sinus arrhythmia: relations to cardiac vagal tone, evolution and biobehavioral functions. *Biological psychology*, 74(2):263–285.
- Grossman, T. and Balakrishnan, R. (2006). The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor’s activation area. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 2006)*, pages 281–290. ACM Press.

- Haag, A., Goronzy, S., Schaich, P., and Williams, J. (2004). Emotion recognition using bio-sensors: first steps towards an automatic system. *Affective Dialogue Systems*, pages 36–48.
- Han, J. N., Stegen, K., De Valck, C., Clément, J., and Van de Woestijne, K. P. (1996). Influence of breathing therapy on complaints, anxiety and breathing pattern in patients with hyperventilation syndrome and anxiety disorders. *Journal of Psychosomatic Research*, 41(5):481–493.
- Hancock, P. A. and Szalma, J. L. (2008). *Performance Under Stress*. Ashgate, Williston, VT, USA.
- Hartmann, T., Toz, E., and Brandon, M. (2010). Just a game? Unjustified virtual violence produces guilt in empathetic players. *Media Psychology*, 13(4):339–363.
- Hazlett, R. L. (2006). Measuring emotional valence during interactive experiences: boys at video game play. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 2006)*, pages 1023–1026. ACM Press.
- Hazlett-Stevens, H. (2008). *Psychological Approaches to Generalized Anxiety Disorder*. Springer, Boston, MA, USA.
- Healey, J. A. and Picard, R. W. (1998). StartleCam: a cybernetic wearable camera. Technical Report 468, MIT Media Lab.
- Healey, J. A. and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166.
- Hernandez, J., Morris, R. R., and Picard, R. W. (2011). Call center stress recognition with person-specific models. In *Affective Computing and Intelligent Interaction*, volume 6974 of *Lecture Notes in Computer Science*, pages 125–134.
- Holloway, E. A. and West, R. J. (2007). Integrated breathing and relaxation training (the Papworth method) for adults with asthma in primary care: a randomised controlled trial. *Thorax*, 62(12):1039–1042.
- Hoover, A. and Muth, E. R. (2004). A real-time index of vagal activity. *International Journal of Human-Computer Interaction*, 17(2):197–209.
- Hunyor, S. N., Henderson, R. J., Lal, S. K. L., Carter, N. L., Kobler, H., Jones, M., Bartrop, R. W., Craig, A., and Mihailidou, A. S. (1997). Placebo-controlled biofeedback blood pressure effect in hypertensive humans. *Hypertension*, 29:1225–1231.
- IJsselsteijn, W., Poels, K., and de Kort, Y. A. W. (2008). The Game Experience Questionnaire: development of a self-report measure to assess player experiences of digital games. FUGA Deliverable D3.3, Eindhoven University of Technology, Eindhoven, The Netherlands.
- IQPuzz (2013). Paced Breathing. <https://play.google.com/store/apps/details?id=com.apps.paced.breathing> (last accessed on 11 June 2013).
- Iyer, A., Cosand, L., Courtney, C., Rizzo, A., and Parsons, T. D. (2009). Considerations for designing response quantification procedures in non-traditional psychophysiological applications. In *Foundations of Augmented Cognition — Neuroergonomics and Operational Neuroscience*, volume 5638 of *Lecture Notes in Computer Science*, pages 479–487. Springer.
- Jaeger, J. A., Echiverri, A., Zoellner, L. A., Post, L., and Feeny, N. C. (2009). Factors associated with choice of exposure therapy for PTSD. *International Journal of Behavioral and Consultation Therapy*, 5(3–4):294–310.
- James, W. (1884). What is an emotion? *Mind*, 9(34):188–205.
- James, W. (1890). *The Principles of Psychology*. Holt, New York, NY, USA.

- Jang, J.-S. R., Sun, C.-T., and Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing*. Prentice Hall, Upper Saddle River, NJ, USA.
- Jasper, H. H. (1958). The ten twenty electrode system of the International Federation. *Electroencephalography and clinical neurophysiology*, 10:371–375.
- Jensen, M. P., Barber, J., Romano, J. M., Hanley, M. A., Raichle, K. A., Molton, I. R., Engel, J. M., Osborne, T. L., Stoelb, B. L., Cardenas, D. D., and Patterson, D. R. (2009). Effects of self-hypnosis training and emg biofeedback relaxation training on chronic pain in persons with spinal-cord injury. *International Journal of Clinical and Experimental Hypnosis*, 57(3):239–268.
- Jones, A. Y. M., Kam, C., Lai, K. W., Lee, H. Y., Chow, H. T., Lau, S. F., Wong, L. M., and He, J. (2003). Changes in heart rate and R-wave amplitude with posture. *Chinese Journal of Physiology*, 46(2):63–69.
- Jönsson, P. (2007). Respiratory sinus arrhythmia as a function of state anxiety in healthy individuals. *International Journal of Psychophysiology*, 63(1):48–54.
- Joseph, C. N., Porta, C., Casucci, G., Casiraghi, N., Maffei, M., Rossi, M., and Bernardi, L. (2005). Slow breathing improves arterial baroreflex sensitivity and decreases blood pressure in essential hypertension. *Hypertension*, 46(4):714–718.
- Kato, P. M. (2010). Video games in health care: closing the gap. *Review of General Psychology*, 14(2):113–121.
- Katsis, C. D., Katertsidis, N. S., and Fotiadis, D. I. (2011). An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders. *Biomedical Signal Processing and Control*, 6(3):261–268.
- Katsis, C. D., Katertsidis, N. S., Ganiatsas, G., and Fotiadis, D. I. (2008). Toward emotion recognition in car-racing drivers: a biosignal processing approach. *IEEE Transactions on Systems, Man, and Cybernetics — Part A: Systems and Humans*, 38(3):502–512.
- Kim, K. H., Bang, S. W., and Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, 42:419–427.
- Kirstein, C. and Muller, H. (1998). Interaction with a projection screen using a camera-tracked laser pointer. In *Proc. 1998 Conference on MultiMedia Modeling (MMM 1998)*, page 191. IEEE Computer Society.
- Konami (2007). Eledees. <http://uk.games.konami-europe.com/game.do?idGame=141> (last accessed on 19 June 2013).
- König, W. A., Bieg, H. J., Schmidt, T., and Reiterer, H. (2007). Position-independent interaction for large high-resolution displays. In *Proc. IADIS International Conference on Interfaces and Human Computer Interaction 2007 (IHCI 2007)*, pages 117–125. IADIS Press.
- Kopin, I. J., Eisenhofer, G., and Goldstein, D. (1988). Sympathoadrenal medullary system and stress. *Advances in Experimental Medicine and Biology*, 245:11–23.
- Kopper, R., Bacim, F., and Bowman, D. A. (2011). Rapid and accurate 3D selection by progressive refinement. In *Proc. 2011 IEEE Symposium on 3D User Interfaces (3DUI 2011)*, pages 67–74. IEEE Computer Society.
- Landelijk Centrum Stressmanagement (2009). Vital-EQ Respiroguide. <http://www.heartintelligence.com/11/index.html> (last accessed on 11 June 2013).
- Landis, C. and Hunt, W. A. (1939). *The Startle Pattern*. Farrar, New York, NY, USA.
- Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372.

- Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). International Affective Picture System (IAPS): Technical manual and affective ratings. NIMH Center for the Study of Emotion and Attention.
- Lang, P. J., Davis, M., and Öhman, A. (2000). Fear and anxiety: animal models and human cognitive psychophysiology. *Journal of Affective Disorders*, 61:137–159.
- Lange, C. (1887). *Ueber Gemuthsbewegungen*, volume 3. Dogma.
- Lazarus, R. S. (1968). Emotions and adaptation: conceptual and empirical relations. *Nebraska Symposium on Motivation*, 16:175–266.
- Lazarus, R. S. (1991). *Emotion and Motivation*. Oxford University Press, New York, NY, USA.
- Lazarus, R. S. and Folkman, S. (1984). *Stress, Appraisal, and Coping*. Springer, New York, NY, USA.
- Liao, W., Zhang, W., Zhu, Z., and Ji, Q. (2005). A decision theoretic model for stress recognition and user assistance. In *Proc. 20th National Conference on Artificial Intelligence (AAAI 2005)*, pages 529–534. AAAI Press.
- Liao, W., Zhang, W., Zhu, Z., Ji, Q., and Gray, W. D. (2006). Toward a decision–theoretic framework for affect recognition and user assistance. *International Journal of Human–Computer Studies*, 64(9):847–873.
- Lim, C. L., Rennie, C., Barry, R. J., Bahramali, H., Lazzaro, I., Manor, B., and Gordon, E. (1997). Decomposing skin conductance into tonic and phasic components. *International Journal of Psychophysiology*, 25(2):97–109.
- Lisetti, C. L. and Nasoz, F. (2004). Using noninvasive wearable computers to recognize human emotions from physiological signals. *Eurasip Journal on Applied Signal Processing*, 2004(11):1672–1687.
- Liu, C., Agrawal, P., Sarkar, N., and Chen, S. (2009). Dynamic difficulty adjustment in computer games through real–time anxiety–based affective feedback. *International Journal of Human–Computer Interaction*, 25(6):506–529.
- Liu, G.-Z., Huang, B.-Y., and Wang, L. (2011). A wearable respiratory biofeedback system based on generalized body sensor network. *Telemedicine and e-Health*, 17(5):348–357.
- Lubar, J. F. (1991). Discourse on the development of EEG diagnostics and biofeedback for attention–deficit/hyperactivity disorders. *Biofeedback and Self-Regulation*, 16(3):201–225.
- Lubar, J. F. and Shouse, M. N. (1976). EEG and behavioral changes in a hyperkinetic child concurrent with training of the sensorimotor rhythm (SMR). *Biofeedback and Self-Regulation*, 1(3):293–306.
- MacKenzie, I. S. and Jusoh, S. (2001). An evaluation of two input devices for remote pointing. *Engineering for Human–Computer Interaction*, pages 235–250.
- MacKenzie, I. S., Kauppinen, T., and Silfverberg, M. (2001). Accuracy measures for evaluating computer pointing devices. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 2001)*, pages 9–16. ACM Press.
- Mamdani, E. H. and Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man–Machine Studies*, 7(1):1–13.
- Mandryk, R. L. and Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human–Computer Studies*, 65(4):329–347.
- Martens, A., Kosloff, S., Greenberg, J., Landau, M. J., and Schmader, T. (2007). Killing begets killing: evidence from a bug–killing paradigm that initial killing fuels subsequent killing. *Personality and Social Psychology Bulletin*, 33:1251–1264.

- Matveyev, S. V. and Göbel, M. (2003). The optical tweezers: multiple-point interaction technique. *Proc. ACM Symposium on Virtual Reality Software and Technology (VRST 2003)*, pages 184–187.
- Mauss, I. B. and Robinson, M. D. (2009). Measures of emotion: a review. *Cognition & Emotion*, 23(2):209–237.
- McMahan, R. P., Ragan, E. D., Leal, A., Beaton, R. J., and Bowman, D. A. (2011). Considerations for the use of commercial video games in controlled experiments. *Entertainment Computing*, 2(1):3–9.
- Meichenbaum, D. H. and Deffenbacher, J. L. (1988). Stress inoculation training. *The Counseling Psychologist*, 16(1):69–90.
- Merletti, R., Knaflitz, M., and De Luca, C. J. (1990). Myoelectric manifestations of fatigue in voluntary and electrically elicited contractions. *Journal of Applied Physiology*, 69(5):1810–1820.
- Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, 7(3):221–237.
- Mine, M. R. (1995). Virtual environment interaction techniques. Technical Report TR95-018, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
- Mitchell, E., Coyle, S., O’Connor, N. E., Diamond, D., and Ward, T. (2010). Breathing feedback system with wearable textile sensors. In *2010 International Conference on Body Sensor Networks (BSN 2010)*, pages 56–61. IEEE Computer Society Conference Publishing Services.
- Morris, M. and Guilak, F. (2009). Mobile Heart Health: project highlight. *IEEE Pervasive Computing*, 8(2):57–61.
- Morris, M. E., Kathawala, Q., Leen, T. K., Gorenstein, E. E., Guilak, F., Labhard, M., and Deleeuw, W. (2010). Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. *Journal of Medical Internet Research*, 12(2):e10.
- Mühlberger, A., Bühlhoff, H. H., Wiedemann, G., and Pauli, P. (2007). Virtual reality for the psychophysiological assessment of phobic fear: responses during virtual tunnel driving. *Psychological Assessment*, 19(3):340–346.
- Mühlberger, A., Wieser, M. J., and Pauli, P. (2008). Darkness-enhanced startle responses in ecologically valid environments: a virtual tunnel driving experiment. *Biological psychology*, 77(1):47–52.
- Mulder, J. D. (1998). Remote object translation methods for immersive virtual environments. In Göbel, M., Landauer, J., Lang, U., and Wapler, M., editors, *Virtual Environments ’98*, Eurographics, pages 80–89. Springer Vienna.
- Muller, M. (2004). Multiple paradigms in affective computing. *Interacting with Computers*, 16(4):759–768.
- Muth, E. R., Moss, J. D., Rosopa, P. J., Salley, J. N., and Walker, A. D. (2012). Respiratory sinus arrhythmia as a measure of cognitive workload. *International Journal of Psychophysiology*, 83(1).
- MyBrainSolutions (2012). MyCalmBeat. <https://www.mybrainsolutions.com/Pages/TrainingThatsFun/GoMobile.aspx> (last accessed on 11 June 2013).
- Myers, B. A., Bhatnagar, R., Nichols, J., Peck, C. H., Kong, D., Miller, R., and Long, A. C. (2002). Interacting at a distance: measuring the performance of laser pointers and other devices. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 2002)*, pages 33–40. ACM Press.
- Myers, K. M. and Davis, M. (2007). Mechanisms of fear extinction. *Molecular Psychiatry*, 12(2):120–150.
- Nacke, L. E. and Lindley, C. A. (2009). Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. *Loading...*, 3(5).

- Nakasone, A., Prendinger, H., and Ishizuka, M. (2005). Emotion recognition from electromyography and skin conductance. In *Proc. 5th International Workshop on Biosignal Interpretation (BSI 2005)*, pages 1–4. Congress Secretariat.
- Nasoz, F., Lisetti, C. L., and Vasilakos, A. V. (2010). Affectively intelligent and adaptive car interfaces. *Information Sciences*, 180(20):3817–3836.
- National Center for Telehealth and Technology (2011a). Breathe2Relax. <http://t2health.org/apps/breathe-2relax#.UbcZsPm-13s> (last accessed on 10 June 2013).
- National Center for Telehealth and Technology (2011b). Tactical Breather. <http://www.t2health.org/apps/tactical-breather> (last accessed on 10 June 2013).
- Nichols, S., Haldane, C., and Wilson, J. R. (2000). Measurement of presence and its consequences in virtual environments. *International Journal of Human–Computer Studies*, 52(3):471–491.
- Oh, J.-Y. and Stuerzlinger, W. (2002). Laser pointers as collaborative pointing devices. *Graphics Interface*, pages 141–150.
- Öhman, A. (2008). Fear and Anxiety. In Lewis, M. and Haviland-Jones, J. M., editors, *Handbook of Emotions*, pages 709–729. The Guilford Press, New York, NY.
- Olsen Jr, D. R. and Nielsen, T. (2001). Laser pointer interaction. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 2001)*, pages 17–22. ACM Press.
- Oshuga, M., Shimono, F., and Genno, H. (2001). Assessment of phasic work stress using autonomic indices. *International Journal of Psychophysiology*, 40:211–220.
- Ouramdane, N., Davesne, F., Otmane, S., and Mallem, M. (2006a). 3D interaction technique to enhance telemanipulation tasks using virtual environment. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006)*, pages 5201–5207. IEEE Computer Society.
- Ouramdane, N., Otmane, S., Davesne, F., and Mallem, M. (2006b). FOLLOW–ME: a new 3D interaction technique based on virtual guides and granularity of interaction. In *Proc. 2006 ACM International Conference on Virtual Reality Continuum and its Applications (VRCIA 2006)*, pages 137–144. ACM Press.
- Pallavicini, F., Cipresso, P., Raspelli, S., Grassi, A., Serino, S., Vigna, C., Triberti, S., Villamira, M., Gaggioli, A., and Riva, G. (2013). Is virtual reality always an effective stressor for exposure treatments? Some insights from a controlled trial. *BMC Psychiatry*, 13(52).
- Pan, J. and Tompkins, W. J. (1985). A real–time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, 32(3):230–236.
- Papez, J. W. (1937). A proposed mechanism of emotion. *Archives of Neurology and Psychiatry*, 38(4):725.
- Parsons, T. D., Kenny, P., Cosand, L., Iyer, A., Courtney, C., and Rizzo, A. A. (2009). A virtual human agent for assessing bias in novice therapists. *Studies in Health Technology and Informatics*, 142:253–258.
- Parsons, T. D., Rizzo, A. A., Courtney, C. G., and Dawson, M. E. (2012). Psychophysiology to assess impact of varying levels of simulation fidelity in a threat environment. *Advances in Human–Computer Interaction*, 2012(5):1–9.
- Pavlovych, A. and Stuerzlinger, W. (2009). The tradeoff between spatial jitter and latency in pointing tasks. In *Proc. of the 1st ACM SIGCHI symposium on Engineering interactive computing systems (EICS 2009)*, pages 187–196, New York, NY, USA. ACM Press.
- Pearl, J. (1985). Bayesian networks: a model of self–activated memory for evidential reasoning. Technical Report R-43, Cognitive System Laboratory, University of California.
- Picard, R. W. (1997). *Affective computing*. MIT Press, Cambridge, MA, USA.

- Picard, R. W. and Goodwin, M. S. (2008). Developing innovative technology for future personalized autism research and treatment. *Autism Advocate*, 50(1):32–39.
- Picard, R. W., Vyzas, E., and Healey, J. A. (2001). Toward machine emotional intelligence analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191.
- Pistoia, F., Sacco, S., and Carolei, A. (2013). Behavioral therapy for chronic migraine. *Current Pain and Headache Reports*, 17(1):304–311.
- Pizzagalli, D. A. (2007). Electroencephalography and high-density electrophysiological source localization. In Cacioppo, J. T., Tassinari, L. G., and Berntson, G. G., editors, *Handbook of Psychophysiology*, pages 56–84. Cambridge University Press, New York, NY, USA.
- Popović, S., Horvat, M., Kukolja, D., Dropuljić, B., and Cosić, K. (2009). Stress inoculation training supported by physiology-driven adaptive virtual reality stimulation. *Studies in Health Technology and Informatics*, 144:50–54.
- Poupyrev, I., Billinghamurst, M., Weghorst, S., and Ichikawa, T. (1996). The Go-Go interaction technique: non-linear mapping for direct manipulation in VR. In *Proc. 9th annual ACM Symposium on User Interface Software and Technology (UIST 1996)*, pages 79–80. ACM Press.
- Radaelli, A., Raco, R., Perfetti, P., Viola, A., Azzellino, A., Signorini, M. G., and Ferrari, A. U. (2004). Effects of slow, controlled breathing on baroreceptor control of heart rate and blood pressure in healthy men. *Journal of Hypertension*, 22(7):1361–1370.
- Rani, P., Sarkar, N., Smith, C. A., and Adams, J. A. (2007). Anxiety-based affective communication for implicit human-machine interaction. *Advanced Engineering Informatics*, 21:323–334.
- Rani, P., Sarkar, N., Smith, C. A., and Kirby, L. D. (2004). Anxiety detecting robotic system towards implicit human-robot collaboration. *Robotica*, 22(1):85–95.
- Ravaja, N., Turpeinen, M., Saari, T., Puttonen, S., and Keltikangas-Järvinen, L. (2008). The psychophysiology of James Bond: phasic emotional responses to violent video game events. *Emotion*, 8(1):114–120.
- Reynolds, S. B. (1984). Biofeedback, relaxation training, and music: homeostasis for coping with stress. *Biofeedback and Self-Regulation*, 9(2):169–179.
- Rezazadeh, I. M., Firoozabadi, M., Hu, H., and Hashemi Golpayegani, M. R. (2012). Co-adaptive and affective human-machine interface for improving training performances of virtual myoelectric forearm prosthesis. *IEEE Transactions on Affective Computing*, 3(3):285–297.
- Riva, G. and Mantovani, F. (2012). From the body to the tools and back: a general framework for presence in mediated interactions. *Interacting with Computers*, 24(4):203–210.
- Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C., and Alcañiz, M. (2007). Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior*, 10(1):45–56.
- Rozmaryn, L. M. (2005). Tendinopathy in the Workplace. In Maffulli, N., Renström, P., and Leadbetter, W. B., editors, *Tendon Injuries: Basic Science and Clinical Medicine*, pages 90–100. Springer-Verlag, London, UK.
- Rule, B. G., Taylor, B. R., and Dobbs, A. R. (1987). Priming effects of heat on aggressive thoughts. *Social Cognition*, 5(2):131–143.
- Rupert, P. A. and Holmes, D. S. (1978). Effects of multiple sessions of true and placebo heart rate biofeedback training on the heart rates and anxiety levels of anxious patients during and following treatment. *Psychophysiology*, 15(6):582–590.

- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.
- Russell, J. A., Weiss, A., and Mendelsohn, G. A. (1989). Affect Grid: a single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, 57(3):493–502.
- Saagara (2011). Universal Breathing: Pranayama. <http://www.saagara.com/apps/pranayama-app> (last accessed on 11 June 2013).
- Sass, S. M., Heller, W., Stewart, J. L., Levin Siltan, R., Edgar, J. C., Fisher, J. E., and Miller, G. A. (2010). Time course of attentional bias in anxiety: emotion and gender specificity. *Psychophysiology*, 47(2):247–259.
- Schachter, S. and Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69(5):379–399.
- Schein, M. H., Gavish, B., Herz, M., Rosner-Kahana, D., Naveh, P., Knishkowsky, B., Zlotnikov, E., Ben-Zvi, N., and Melmed, R. N. (2001). Treating hypertension with a device that slows and regularises breathing: a randomised, double-blind controlled study. *Journal of human hypertension*, 15(4):271–278.
- Scheirer, J., Fernandez, R., Klein, J., and Picard, R. W. (2002). Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers*, 14(2):93–118.
- Scherer, K. R. (1984). On the nature and the function of emotion: a component process approach. In Scherer, K. R. and Ekman, P., editors, *Approaches to Emotion*, pages 293–317. Erlbaum, Hillsdale, NJ, USA.
- Scheumann, D. W. (2007). *The Balanced Body: A Guide to Deep Tissue And Neuromuscular Therapy*. Lippincott Williams & Wilkins, Baltimore, MD, USA, 3rd edition.
- Sestir, M. A. and Bartholow, B. D. (2010). Violent and nonviolent video games produce opposing effects on aggressive and prosocial outcomes. *Journal of Experimental Social Psychology*, 46(6):934–942.
- Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):1–22.
- Shizuki, B., Hisamatsu, T., Takahashi, S., and Tanaka, J. (2006). Laser pointer interaction techniques using peripheral areas of screens. In *Proc. of the Working Conference on Advanced Visual Interfaces (AVI 2006)*, pages 95–98. ACM Press.
- Shusterman, V. and Barnea, O. (2005). Sympathetic nervous system activity in stress and biofeedback relaxation. *IEEE Engineering in Medicine and Biology Magazine*, 24(2):52–57.
- Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., Garau, M., Brogni, A., and Friedman, D. (2006). Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence*, 15(5):553–569.
- Sommerich, C. M., Marras, W. S., and Karwowski, W. (2006). Work-Related Upper Extremity Musculoskeletal Disorders. In Salvendy, G., editor, *Handbook of Human Factors and Ergonomics*, pages 855–888. John Wiley & Sons, Hoboken, NJ, USA.
- Sylvers, P., Lilienfeld, S. O., and LaPrairie, J. L. (2011). Differences between trait fear and trait anxiety: implications for psychopathology. *Clinical Psychology Review*, 31(1):122–137.
- 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. *Circulation*, 93:1043–1065.
- Tassinari, L. G., Cacioppo, J. T., and Vanman, E. J. (2007). The Skeletomotor System: Surface Electromyography. In *Handbook of Psychophysiology*, pages 267–302. Cambridge University Press.

- Taveira, A. D. and Smith, M. J. (2006). Social and Organizational Foundations of Ergonomics. In Salvendy, G., editor, *Handbook of Human Factors and Ergonomics*, pages 269–291. Wiley, Hoboken, NJ, USA.
- Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality*, 35(2):297–310.
- Teather, R. J., Pavlovych, A., Stuerzlinger, W., and MacKenzie, I. S. (2009). Effects of tracking technology, latency, and spatial jitter on object movement. In *Proc. IEEE Symposium on 3D User Interfaces (3DUI 2009)*, pages 43–50. IEEE Computer Society.
- The Association for Applied Psychophysiology and Biofeedback (2011). About Biofeedback. <http://www.aapb.org/i4a/pages/index.cfm?pageid=3463> (last accessed on 26 June 2013).
- Thought Technology (2013). Biograph Infiniti. <http://www.thoughttechnology.com/bioinf.htm> (last accessed on 17 December 2013).
- TotalAwake (2012). TotalAwake. <http://totalawake.com/index.php> (last accessed on 11 June 2013).
- Tsai, P.-S., Chang, N.-C., Chang, W.-Y., Lee, P.-H., and Wang, M.-Y. (2007). Blood pressure biofeedback exerts intermediate-term effects on blood pressure and pressure reactivity in individuals with mild hypertension: a randomized controlled study. *The Journal of Alternative and Complementary Medicine*, 13(5):547–554.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison Wesley, Reading, MA, USA.
- Unity Technologies (2013). Unity. <http://unity3d.com/> (last accessed on 17 December 2013).
- Van Den Broek, E. L. (2011). *Affective Signal Processing (ASP)*. PhD thesis, University of Twente.
- Van Den Wittenboer, G., Van Der Wolf, K., and Van Dixhoorn, J. (2003). Respiratory variability and psychological well-being in schoolchildren. *Behavior Modification*, 27(5):653–670.
- Van Diest, I., Thayer, J. F., Vandeputte, B., Van de Woestijne, K. P., and Van den Bergh, O. (2006). Anxiety and respiratory variability. *Physiology & Behavior*, 89(2):189–195.
- Vasudeva, S., Claggett, A. L., Tietjen, G. E., and McGrady, A. V. (2003). Biofeedback-assisted relaxation in migraine headache: relationship to cerebral blood flow velocity in the middle cerebral artery. *Headache*, 43(3):245–250.
- Vila, J., Guerra, P., Muñoz, M. Á., Vico, C., Viedma-del Jesús, M. I., Delgado, L. C., Perakakis, P., Kley, E., Mata, J. L., and Rodríguez, S. (2007). Cardiac defense: from attention to action. *International Journal of Psychophysiology*, 66(3):169–182.
- Villani, D., Riva, F., and Riva, G. (2007). New technologies for relaxation: the role of presence. *International Journal of Stress Management*, 14(3):260–274.
- Villani, D. and Riva, G. (2012). Does Interactive Media Enhance the Management of Stress? Suggestions from a Controlled Study. *Cyberpsychology, Behavior, and Social Networking*, 15(1):24–30.
- Wahlstrom, J. (2005). Ergonomics, musculoskeletal disorders and computer work. *Occupational Medicine*, 55(3):168–176.
- Ward, R. D. and Marsden, P. H. (2004). Affective computing: problems, reactions and intentions. *Interacting with Computers*, 16(4):707–713.
- Wells, G. L. and Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, 25(9):1115–1125.

- Wu, D., Courtney, C. G., Lance, B. J., Narayanan, S. S., Dawson, M. E., Oie, K. S., and Parsons, T. D. (2010). Optimal arousal identification and classification for affective computing using physiological signals: virtual reality Stroop task. *IEEE Transactions on Affective Computing*, 1(2):109–118.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353.
- Zaki, J., Davis, J. I., and Ochsner, K. N. (2012). Overlapping activity in anterior insula during interoception and emotional experience. *Neuroimage*, 62(1):493–499.
- Zhai, J. and Barreto, A. (2006). Stress detection in computer users through non-invasive monitoring of physiological signals. *Biomedical Sciences Instrumentation*, 42:495–500.
- Zong, W., Heldt, T., Moody, G., and Mark, R. (2003). An open-source algorithm to detect onset of arterial blood pressure pulses. *Computers in Cardiology*, 30:259–262.