NOTICE: This is the author's version of a work that was accepted for publication in the **International Journal of Human-Computer Studies**. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms, may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in the **International Journal of Human-Computer Studies**, Volume 72, Issue 8-9, August-September 2014, DOI: http://dx.doi.org/10.1016/j.ijhcs.2014.01.007.

Please cite as:
Chittaro L., Sioni R. Affective computing vs. affective placebo: Study of a biofeedback-controlled game for relaxation training, *International Journal of Human-Computer Studies, 72(8-9)*, 2014, pp. 663-673.

Affective Computing vs. Affective Placebo: Study of a biofeedback-controlled game for relaxation training

Luca Chittaro and Riccardo Sioni

Human-Computer Interaction Lab
Department of Mathematics and Computer Science
University of Udine
Via delle Scienze 206, 33100, Udine, Italy

Abstract

Relaxation training is an application of affective computing with important implications for health and wellness. After detecting user's affective state through physiological sensors, a relaxation training application can provide the user with explicit feedback about his/her detected affective state. This process (biofeedback) can enable an individual to learn over time how to change his/her physiological activity for the purposes of improving health and performance. In this paper, we provide three contributions to the field of affective computing for health and wellness. First, we propose a novel application for relaxation training that combines ideas from affective computing and games. The game detects user's level of stress and uses it to influence the affective state and the behavior of a 3D virtual character as a form of embodied feedback. Second, we compare two algorithms for stress detection which follow two different approaches in the affective computing literature: a more practical and less costly approach that uses a single physiological sensor (skin conductance), and a potentially more accurate approach that uses four sensors (skin conductance, heart rate, muscle activity of corrugator supercilii and zygomaticus major). Third, as the central motivation of our research, we aim to improve the traditional methodology employed for comparisons in affective computing studies. To do so, we add to the study 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 obtained results show that only the feedback presented by the single-sensor algorithm was perceived as significantly more accurate than the placebo. If the placebo condition were not included in the study, the effectiveness of the two algorithms would have instead appeared similar. This outcome highlights the importance of using more thorough methodologies in future affective computing studies.

Keywords: affective computing, placebo effect, evaluation methods, biofeedback, embodiment, relaxation, training, virtual characters

1. Introduction

Affective computing systems aim at detecting user's affective state and change their behavior based on that information (Picard, 1997). Relaxation training is one of the applications of affective computing which have important implications for health and wellness. After detecting user's affective state using physiological sensors, the application can provide the user with explicit feedback about his/her detected affective state. This process (*biofeedback*) can enable an individual to learn over time how to change his/her physiological activity for the purposes of improving health and performance (Association for Applied Psychophysiology and Biofeedback, 2011).

More specifically, biofeedback applications measure users' physiological signals like *electrodermal activity* (EDA) and *heart rate* (HR), and "feed back" this information to the user who, over time, can learn to consciously control his/her physiological processes and support the desired physiological changes. Biofeedback is employed to improve individual well-being in the treatment of medical and medical-related conditions, such as pain (Jensen et al., 2009), incontinence (Enck, 1993), migraine (Fentress, Masek, Mehegan, & Benson, 1986). Moreover, it can be used to reduce stress-related symptoms (Bouchard, Bernier, Boivin, Morin, & Robillard, 2012) and enhance personal well-being (Chandler, Bodenhamer-Davis, Holden, Evenson, & Bratton, 2001).

Many biofeedback systems in the literature still employ very simple acoustic or visual cues to provide users with information 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 user's arousal. However, in recent years, some authors started to employ richer stimuli such as virtual reality and video games to convey biofeedback to users, e.g. (Bersak et al., 2001; Bouchard et al., 2012). An advantage of this approach is that immersive and realistic *virtual environments* (VEs) draw more of the users' attention and can make the feedback more effective (Bersak et al., 2001).

The goal of this paper is threefold. First, we propose a novel biofeedback application for relaxation training that combines ideas from affective computing and games. We employ a 3D VE to present users with real-life simulated scenarios in which they have to focus on maintaining relaxation when faced with stressors. The level of stress detected through physiological signals determines the facial expressions as well as the behavior of a virtual character that represents the player in the VE. This relation between user's affective state and the virtual character provides users with an embodied visual and acoustic feedback about their state of relaxation. To detect users' level of stress, the game can employ up to four physiological sensors: EDA, HR, and two *electromyography* (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 they use. To relate our research to well-known approaches proposed in the affective computing literature, one of the stress detection algorithms we implemented is inspired by Healey and Picard (1998) while the other follows Mandryk and Atkins (2007). In particular, the first approach exploits only the EDA sensor while the second considers all the four physiological sensors mentioned above to compute the stress index for the biofeedback application. We are interested in comparing these two approaches in the context of biofeedback-based relaxation training because they may have different advantages: a single-sensor approach is more practical and less costly, but the use of multiple physiological sensors may improve the accuracy of stress detection by taking into account different bodily responses.

Third, as the central motivation of our research, we aim to improve the traditional methodology employed for comparisons in affective computing studies. To do so, we add to the study 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. 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 actually contain it). In affective computing studies, one cannot rule out the possibility that the suggestibility of participants could alter their perception of the feedback provided by the application, leading them to believe that a real biofeedback system is in place although the provided feedback is actually sham. 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 affective computing systems, 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, the affective computing literature has not yet explored this way of evaluating systems.

The paper is organized as follows. In Section 2, we review the literature on biofeedback applications and evaluation of affective computing systems. Section 3 describes in detail the proposed biofeedback game. Then, Sections 4 and 5 describe in detail our experiment and its results, while Section 6 critically discuss the results. Finally, Section 7 presents conclusions and future work.

2. Related Work

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 follows:

> 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. (Association for Applied Psychophysiology and Biofeedback, 2011)

The use of biofeedback in relaxation training and stress treatment is widely described in the literature. It aims to improve trainees' health and wellness by increasing their ability to relax, and to make them learn how to better cope with stress. Indeed, high levels of arousal are often hallmarks of anxiety disorders and stress (Chandler et al., 2001), which can threaten 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, Sacco, & Carolei, 2013) and increase feelings of well-being (Chandler et al, 2001; Critchley et al., 2001; Pistoia et al., 2013).

Biofeedback-based relaxation training is also used in the treatment of some medical conditions such as chronic pain (Jensen et al., 2009), fibromyalgia (Buckelew et al., 2005), migraine (Fentress et al., 1986; Pistoia et al., 2013; Vasudeva, Claggett, Tietjen, & McGrady, 2003), and hypertension (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., 2005, 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 considered, 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 cost of biofeedback equipment is justifiable when simpler and less expensive alternatives are available. For example, *relaxation response* (Benson, 1975), i.e., a set of coordinated 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) underline 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. For example, this issue is apparent in studies that apply biofeedback to 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 provide short term improvements in numerous functional disorders of the gastrointestinal tract, especially those related to the lower part of the gut (Bassotti & Whitehead, 1997), but more research is instead needed about the long-term effects of biofeedback-based treatments (Enck, 1993).

Attention deficit-hyperactivity disorder (ADHD) is another common target for biofeedback-based treatments, as originally proposed by Lubar & Shouse (1976). The review by Lubar (1991) shows that biofeedback training, although time consuming, can lead to significant improvements in ADHD, while the meta-analysis by Arns, de Riddler, Strehl, Breteler, and Coenen (2009) confirms that it can be regarded as clinically meaningful and its clinical effects seem to be stable and might improve further with time.

To derive physiological indices of arousal and stress, the biofeedback literature often exploits EDA and facial EMG, e.g. (Bouchard et al., 2012; Critchley, Melmed, Featherstone, Mathias and Dolan, 2001; Critchley, Wiens, Rotshtein, Öhman, & Dolan, 2004; Jensen et al., 2009; Reynolds, 1984). Peripheral temperature and cardiac parameters such as HR and *heart rate variability* (HRV) are also employed, e.g. (Bouchard et al., 2012; Shusterman & Barnea, 2005). In some cases, EMG is used as a direct measure for the specific physiological phenomena under treatment. For example, Foster (2004) uses EMG from the masseter muscle (located in the jaw) for the biofeedback-based treatment of chronic nocturnal bruxism, and 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 ADHD, employ the electrical activity of the brain (EEG) to infer the affective state of the patient.

The feedback provided to the user by most biofeedback systems cited above is very basic. Some systems use simple audio stimuli, e.g. a repeating tone (Foster, 2004; Jensen et al., 2009) that decreases in frequency as the EMG activity of masseter and frontalis muscles decreases , indicating a lower level of arousal and stress (Jensen et al., 2009) or a diminishing intensity of the disorder under treatment (Foster, 2004). In other systems, this very basic feedback is visual, e.g. showing the signal graph of one or more physiological signals such as frontal EMG (Fentress et al., 1986) or a thermometer whose temperature is positively correlated to EDA level so that higher arousal is represented by higher temperature (Critchley et al., 2001; Critchley, Melmed, Featherstone, Mathias and Dolan; 2002).

The use of much richer feedback based on realistic 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 organized as a racing game: 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 that employ biofeedback have been considered 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 essentially based on providing information and teaching techniques in a classroom, and highlight instead 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. Dekker and Champion (2007) employed similar feedback cues, but for entertainment purposes, enhancing gameplay 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 gameplay shown by novice and experienced players, and different factors (such as room illumination or the use of a computer with a small screen) 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). Such systems 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 (Hunyor et al., 1997; Rupert & Holmes, 1978; Tsai, Chang, Chang, Lee, & Wang, 2007). Such experiments often led to relevant results, for example, Hunyor et al. (1997) observed that their biofeedback treatment of hypertension was successful in lowering systolic blood pressure using a real as well as a placebo feedback signal.

The affective computing literature usually assesses the accuracy of emotion detection through confusion matrices (Healey & Picard, 2005; Zhai & Barreto, 2006), mean successful recognition (Liu, Agrawal, Sarkar, & Chen, 2009; Nasoz, Lisetti, & Vasilakos, 2010; Wu et al., 2010), or mean recognition error (Rani, Sarkar, Smith, & Adams, 2007). Comparisons among multiple techniques for emotion detection, and in particular stress and anxiety detection, are also common, e.g. (Katsis et al., 2008; Lisetti & Nasoz, 2004; Zhai & Barreto, 2006). In these cases, control conditions are sometimes employed (Nasoz et al., 2010; Rezazadeh, Firoozabadi, Hu, & Hashemi Golpayegani, 2012), but to the best of our knowledge no comparisons with a placebo condition have been performed. This is acknowledged in (Picard & Goodwin, 2008) which explicitly advocates the consideration of the placebo effect in experimental designs that evaluate the efficacy of affective computing systems.

3. The Biofeedback Game

By combining ideas from affective computing and games, we propose a novel application for relaxation training based on biofeedback. Our 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 & Riva, 2008; Villani & Riva, 2012). Unlike biofeedback systems that employ very simple feedback (see Section 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, 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 that represents the player in the VE, can be a key element for the efficacy of VE-based relaxation treatments (Pallavicini et al., 2013; Riva et al., 2007; Villani, Riva, & Riva, 2007). In particular, the more intense is the connection, the greater is the player's sense of being present inside the VE (Riva & 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. Different stressors can affect the user (and the virtual character that displays his/her affective state) inside each VE. 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). If the player keeps his/her stress level low, the virtual character walks along the path to the emergency exit. The character gets slower and slower as the player's level of stress increases, eventually coming to a halt and 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 long and 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.

Stressors are 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 *stress inoculation training* (SIT).

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 paper, the virtual character is called Andrew and the introductory screen informs the player that Andrew has just been hired by a big company and asked to write a business plan as a first assignment. The level shows Andrew working in front of a PC, inside a typical office environment.

The goal of the player 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. The score is shown at the end of the level, which lasts 3 minutes. To complete the assignment, Andrew has to remain relaxed and work continuously at the PC, ignoring a distracting stressor, i.e., a phone that occasionally rings on the desk. In particular, we have set this stressor to occur 3 times during the level: 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 stressor respectively start 20 s, 50 s, and 90 s after the beginning of the level. The situation reproduced by the level is common in an office context, and thus should be easily and quickly understood by 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. If the user is calm and relaxed, then Andrew focuses on progressing in the task and smiles. As user's stress level increases, Andrew becomes more and more unsettled, less and less focused on the work, and shows discontentment through facial expressions and mumblings of growing intensity. When the phone rings, the virtual camera slightly zooms out to include in the current view the phone on the 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 mouse and keyboard (Figure 1a and 1b), and watching the PC screen with a smile on its face (Figure 1c);
- *State 2*: the character starts to show some signs of uneasiness. Andrew continues to work at its task (Figure 1d and 1e), but sometimes moves its chest towards the screen (Figure 1d), mumbles, and shakes its head; it starts to frown, and its smile starts to fade (Figure 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 1g). Andrew tries to continue working as in Figure 1a and 1b, but sometimes stops for a few seconds, reclines back and puts its hands on the legs (Figure 1f);
- *State 4*: the character cannot focus on its task anymore. Andrew stops working (Figure 1h); the mumblings get increasingly loud, and its face becomes tense and angry (Figure 1i). Sometimes, the character raises its hands (Figure 2a and 2b), and throws a punch on the table (Figure 2c);
- *State 5*: the character reaches its maximum level of stress. Its face becomes angrier (figure 1k), it raises its hands (Figure 1j) and throws a sequence of punches on the table (Figures 2b and 2c); its mumblings turn into shouts.

Figure 1. Screenshots of the virtual character in its five different states.

a b c

Figure 2. Screenshots of the virtual character raising its hands and punching the table.

We built the animations employed in the five states 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 animations were instead modeled by key framing, and were slightly emphasized 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 state change, the previous behavioral animation can seamlessly blend into the subsequent 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 their progression 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-s video of each state, and showed the five videos to each participant following a Latin square design. Then, we asked participants to rank the videos in increasing order of stress displayed by the character. All participants were able to correctly rank the five states without making any errors.

3.2. Physiological Measures

We acquire users' raw physiological signals in real time at 2048 Hz through a Procomp Infiniti device. From them, we calculate a stress index, sending it to the game at an 8 Hz frequency. This section introduces the four physiological signals we acquire, while Section 3.3 illustrates how we compute the stress index.

3.2.1. Electrodermal activity.

EDA is a measure of the electrical conductivity of the skin surface (Andreassi, 2007) and is one of the most frequently employed physiological signals for stress detection. EDA changes 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 regulated exclusively by the *sympathetic nervous system* (SNS) (Boucsein, 2006). The SNS controls those activities that are mobilized during emergency and stress situations (*fight-or-flight* response). In situations that evoke high level of arousal, an increase in the activity of the sweat glands can be observed (Boucsein, 2006), leading to what is called *emotional sweating*. An increase in sweating causes a sharp increase of skin conductance, that can be recorded after about 1 s from the event (skin conductance *response*, the *phasic* component of the EDA signal). If no other stressful stimulus is provided, the EDA signal then slowly decreases, reaching after about 4 s a basal value (skin conductance *level*, representing the *tonic* component of the signal) that is strongly related to the person's overall level of arousal.

Emotional sweating is observed mainly on palmar and plantar sites. Therefore, EDA is usually measured on the palms of the hands or on the fingers, with a current (imperceptible to users) applied between a pair of electrodes placed on two adjacent fingers (Andreassi, 2007). There is normally no need for a pretreatment (i.e., skin cleaning) of the sites used for EDA recording.

As reported in the literature, EDA is a suitable physiological signal in the measurement of arousal and stress. In particular, EDA is regarded as a sensitive and valid indicator for the lower arousal range, reflecting small, mostly cognitively determined, variations in arousal, and has been widely employed in the psychophysiology and affective computing literature for this purpose, e.g. (Fleureau, Guillotel, & Huynh-Thu, 2012; Friedman, Suji, & Slater, 2007; Mandryk & Atkins, 2007; Rani et al., 2007; Scheirer, Fernandez, Klein, & Picard, 2002).

3.2.2. Blood volume pulse and heart rate.

The blood volume pulse (BVP) signal is an indication of the amount of blood flowing into the peripheral vessels, like the ones in the fingers, which is generally measured through a *photoplethysmograph* (PPG). The PPG, usually placed on the distal phalanx of the index finger, generates a subtle IR light that is reflected by the capillaries under the skin by an amount that is proportional to the quantity of blood flowing through them. The signal shows a regular pattern made of peaks and valleys, from which various physiological features can be extracted. HR is the frequency of the heartbeat measured in beats per minute (bpm), calculated from the BVP signal by counting the number of peaks per minute, and is one of the most common physiological parameters that can be employed in stress detection. Increases in HR are generally related to emotional activation (as in EDA), and have been used in the literature as a correlate of arousal and stress (Katsis et al., 2011; Lisetti & Nasoz, 2004; Mandryk & Atkins, 2007; Nasoz et al., 2010; Slater et al., 2006). During events evoking high levels of arousal, the heart pumps blood at higher frequency, leading to an increase of HR. Compared to EDA, HR is well suited as an indicator for the higher arousal range and for pronounced and often somatically determined arousal processes (Boucsein, 2006).

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. HR can be also measured with sensors that are more robust against motion artifacts, such as electrocardiogram (ECG) electrodes. These electrodes, however, must be placed in direct contact with the skin of the user's chest, which is less practical and may also be uncomfortable for some users. In the literature, when BVP is preferred to ECG, researchers typically ask participants to avoid movements of the hand with the PPG during the experiment.

3.2.3. Facial electromyography.

The activity of specific facial muscles is strongly related to positive or negative emotional valence (Andreassi, 2007). In particular, the zygomaticus major (located in the cheek) and the corrugator supercilii (located in the forehead) help to distinguish between positively- and negatively-valenced high-arousal emotions (excitement vs. stress), as reported by many studies, e.g. (Fleureau et al., 2012; Liu et al., 2009; Mandryk & Atkins, 2007). These connections are not surprising since zygomaticus muscles are responsible for smiling, while corrugator muscles are responsible for frowning.

The contractions of each muscle fiber generate a small electrical discharge that can be measured through an electrode placed over the skin above the muscle. When the muscle is under load, the electrodes record a signal composed by a series of discharges. This electrical activity is called *surface* EMG (SEMG), to distinguish it from the EMG measured under the skin directly from the muscle. The electrodes employed for facial muscle activity measurement 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 the affective state of the user.

Typically, the setup required to measure the SEMG signal of a muscle consists of two active electrodes placed along the long axis of the muscle, and an inactive electrode (ground) (Andreassi, 2007). To better detect the electrical activity of the underlying muscle, skin cleaning is required, e.g., with alcohol. The recorded signal is typically (i) filtered to remove components under 20-50 Hz and over 500 Hz, thus keeping only the components related to muscle contraction (Andreassi, 2007), (ii) rectified (the raw EMG is a bipolar signal) using a root mean square (RMS) algorithm, and (iii) smoothed using a moving window, to reduce the impact of motion artifacts.

3.3. Users' Stress Detection

To detect users' stress, our study tests two different algorithms. The first employs only the EDA signal, while the second integrates different physiological signals. The implemented algorithms respectively derive from those presented in (Healey & Picard, 1998) and (Mandryk & Atkins, 2007). We have chosen these two algorithms because they are well-known and representative of two categories of stress detection algorithms: a more practical and less costly approach that uses a single sensor (EDA), and a potentially more accurate approach that uses four sensors (EDA, HR, EMG of corrugator supercilii, EMG of zygomaticus major). The EDA-based algorithm by Healey and Picard was originally developed to detect users' SCRs through a wearable computer, while the algorithm by Mandryk and Atkins was developed to detect users' arousal and valence state as well as the level of five discrete emotions (fun, challenge, boredom, frustration, and excitement).

In implementing the multi-sensor algorithm, we have followed the description provided in (Mandryk, 2005; Mandryk & Atkins, 2007) as closely as possible. In implementing the EDA-only algorithm, we generally followed the approach described in (Healey & 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 his/her physiological measurements, and how the stress index is instead generated pseudo-randomly in the placebo condition of our study.

3.3.1. 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 a stressor affects the user, 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. The temporary baseline and the stress index are managed as follows:

- 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 increase phase, when the EDA exceeds for the first time a 0.05 μS threshold over the temporary baseline value, the stress index is incremented by 1. The 0.05 μS threshold is recommended by (Boucsein, 2006) to distinguish a skin conductance response from signal fluctuations. If the EDA signal keeps 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 3 shows an example of use of the EDA-only algorithm.

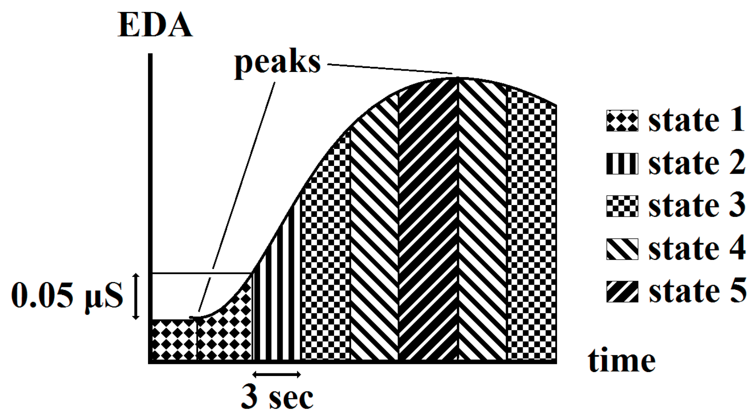

Figure 3. An example of use of the EDA-only algorithm.

This algorithm is simple to implement and 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 the 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 & Picard, 1998), the algorithm does not decompose the EDA signal into phasic and tonic components (see Section 3.2.1), 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.

3.3.2. Multi-sensor algorithm.

The multi-sensor algorithm operates a data fusion of four physiological signals (EDA, HR, EMG of zygomaticus major, EMG of corrugator supercilii) to determine the stress index, following the approach proposed in (Mandryk & Atkins, 2007). 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. As in the EDA-only algorithm, the multi-sensor algorithm does not decompose the EDA signal into phasic and tonic components (see Section 3.2.1).

The multi-sensor algorithm respectively exploits (i) EDA and HR to derive an index of arousal, and (ii) the activities of the two facial muscles to distinguish between positive and negative high-arousal affective states. It maps a negative high-arousal state into a high level of stress, and a positive low-arousal state into a low level of stress. The algorithm filters in real time the four physiological signals as described in the following:

- EDA is filtered by applying a 5 s moving window over the signal (Mandryk & Atkins, 2007);
- To compute HR from BVP, we calculate the *inter-beat interval* (IBI) in milliseconds, i.e., the temporal distance between two heartbeats, by implementing a peak detection algorithm described in (Zong, Heldt, Moody, & Mark, 2003). Since users' movements can easily corrupt the signal from the PPG sensor, the algorithm discards IBI values that can be labeled as motion artifacts, because they are related to HR values that are too high or too low. More precisely, considering the task carried out in our study, we discard IBI values that are lower than 350 ms and higher than 1500 ms, which represent HR values respectively higher than 171 bpm and lower than 40 bpm;
- The EMG data from the corrugator supercilii and the zygomaticus major muscles is rectified by applying a RMS function; the rectified data is then filtered using a 1 s moving window to remove motion artifacts.

The values from the four sensors are compared to a 1-min baseline recorded before the play session, to minimize physiological differences among participants, and then standardized using a t-transformation into a 0-100 range, where 50 corresponds to the mean value of the baseline recording. The t-values from the four physiological sensors are then merged into a single stress index using fuzzy logic (Zadeh, 1965) in a two-step approach that employs the 37 of the 111 fuzzy rules proposed by (Mandryk & Atkins, 2007) which merge:

- HR and EDA t-values into an arousal index in the 0-100 range;
- Zygomaticus and corrugator EMG t-values into a valence index in the 0-100 range;
- Arousal index and valence index into the stress index in the 0-100 range.

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

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

3.3.3. Placebo.

In the placebo condition, the stress index is generated without taking into account user's current physiological measures. Since a completely random assignment would make the overall behavior of the character look implausible and unrealistic, we resorted to a pseudo-random approach for creating sham biofeedback. The process works as follows: (i) we take a physiological data recording of a previous game session, (ii) we pick at random one of the two previously described stress detection algorithms, (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.

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.

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.

4.2. Materials

The game was run on a PC, and displayed in full-screen mode on a 30'', 2560 x 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.

We recorded and processed participants' physiological data on a second PC. To record participants' physiological data, we employed four 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*, placed over the distal phalanx of the index finger.

Physiological data was recorded at 2048 Hz with a Procomp Infinity encoder, and processed in real-time with the two algorithms for stress detection described in Section 3. The calculated stress index was sent through a LAN connection to the PC running the biofeedback game. We assessed the latency of the LAN connection, and the time required to receive the stress index by the game was about 1 ms, which is too short to be perceivable by participants.

To record participants' subjective opinions, we employed a questionnaire that assessed:

- *Perceived quality of the biofeedback*. To assess this aspect, after trying 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 = 0.72);
- *Difficulty of relaxation training*. To assess this aspect, after trying 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 = 0.77).

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 data. They were asked to try the three versions of the game level described in Section 3.1. They were told that their goal was just to relax as much as possible, because the state and ability to focus on its work of Andrew was going to reflect their own relaxation state. Participants were also informed that the whole experiment would last for about 45 minutes. With respect to the description provided in Section 3.1, during the experiment we disabled the score display at the end of the level, and we did not mention scores to participants.

After the introductory briefing, all participants consented to participate in the experiment. They were seated, the skin of their forehead and cheeks was cleaned using a pad of cotton wool and alcohol, and the physiological sensors described in Section 4.2 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. 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 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 sensors. After each game session, participants filled the questionnaire for assessing quality of biofeedback and difficulty of relaxation for the game session just tried.

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

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 probably likely to be small.

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 showed moderate correlations between the two dependent variables (the higher the perceived quality of biofeedback, the smaller the difficulty of relaxation) for the the multi-sensor condition ($\rho = -0.41$, $p < 0.05$) and the EDA-only condition ($\rho = -0.43$, $p < 0.05$) .

The 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 ($p = 0.06$), or the multi-sensor and the placebo conditions ($p = 0.94$). Figure 4 shows the mean values of the feedback score.

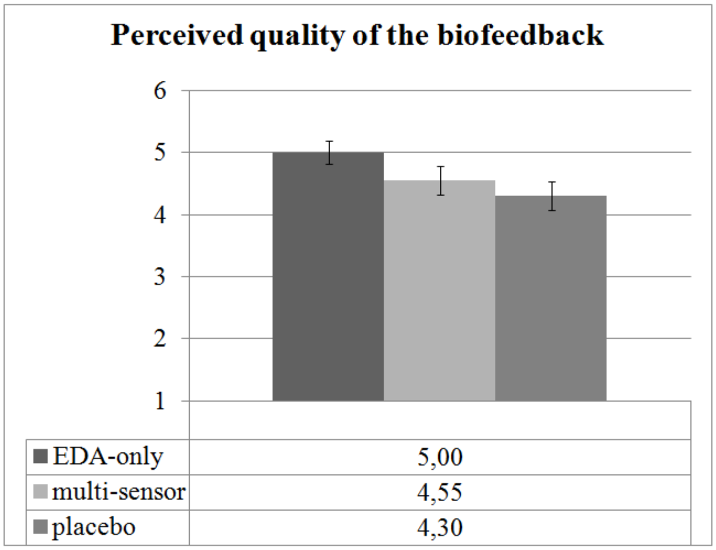

Figure 4. Mean values of the perceived quality of the biofeedback score. Error bars indicate standard error of the mean.

The differences (Figure 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 ($p = 0.63$).

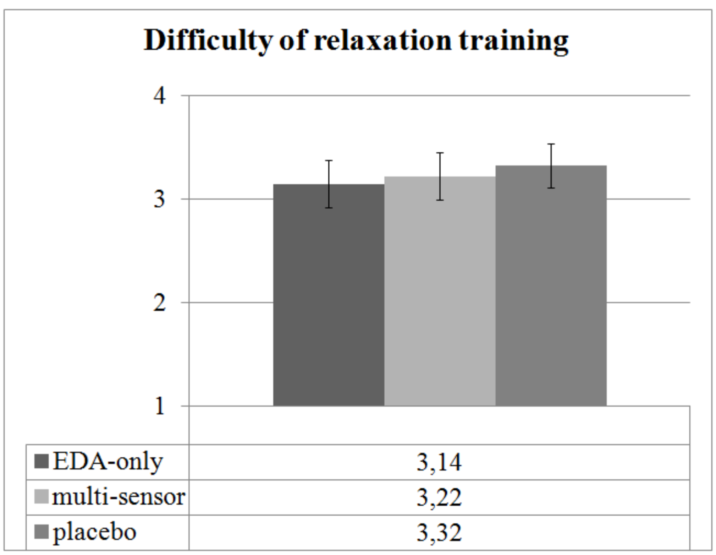

Figure 5. Mean values of the difficulty of relaxation training score. Error bars indicate standard error of the mean.

6. Discussion

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

As expected, the EDA-only technique performed significantly better than the placebo. Contrary to our expectations, the multi-sensor technique did not obtain better ratings with respect to EDA-only. This could 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 3.2.1, 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 3.2.2).

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 auditory (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 most of those stimuli are not pleasant.

However, the very surprising result was that the multi-sensor technique lacked significant differences 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 & Atkins, 2007), which were originally 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 seems to suggest that inferring increases or decreases of the arousal level from the slope of the EDA signal could lead to an higher accuracy in arousal detection with respect to placebo, unlike using a unique baseline value calculated before the 3-min gaming session, as the multi-sensor algorithm did. 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 in EDA (i.e., greater than 0.05 μS) 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 have an effect until the signal value has increased enough to exceed the baseline, which may require time. A similar situation occurs if a player who, after increasing his/her arousal level for a long time, starts relaxing. With the EDA-only algorithm, these EDA increases and decreases can have an immediate effect on the state of the character.

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 technique, 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. The EDA-only condition was able to feed back 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.

In general, our findings highlight the importance of using more thorough methodologies in future affective computing studies, by including placebo conditions to determine if the (often complex and costly) affective computing systems are actually playing a significant role in determining the effectiveness of the feedback. In affective computing studies, methodologies employing instruments and measures such as confusion matrices (Healey & Picard, 2005; Zhai & Barreto, 2006) or mean successful recognition (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 affective computing system 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 added, 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 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 & 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 4.3) were subtracted from the physiological data recorded during the 3-min play session to account for individual differences. 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 partly expected as pointed out in Section 4.4. 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., 2007; Foster, 2004; Jensen et al., 2009). However, the correlation between perceived quality of feedback and difficulty of relaxation training reported in the Results section is promising.

7. Conclusions

This paper has proposed a novel biofeedback system for relaxation training, which was tested with two different stress detection algorithms (single- and multi-sensor) and with sham biofeedback (placebo). Results of the evaluation show that only the feedback produced by the (more convenient and less costly) single-sensor solution 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. This outcome highlights the importance of using more thorough methodologies in future affective computing studies, by including placebo conditions.

The next phase in our research will concern the design of a longitudinal study. While a different quality of feedback can be noticed immediately by users as we have seen in the present study, a longitudinal study would allow one to assess further possible advantages of receiving 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 on participant's health. To better test the capabilities of the multi-sensor approach, we will design new levels for the biofeedback game to present users with auditory and visual cues that can elicit positive emotions, instead of mainly focusing on stressful and negatively-valenced stimuli. This might allow the multi-sensor algorithm to gain an advantage over the EDA-only algorithm, because it could be possibly able to distinguish between affective 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 affective states that could optimize health and well-being (Fredrickson, 2000).

References

Andreassi, J. L. (2007). *Psychophysiology* (5th ed.). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Arns, M., de Ridder, S., Strehl, U., Breteler, M., & 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.

Association for Applied Psychophysiology and Biofeedback. (2011). *About Biofeedback*. aapb.org. Retrieved on April 12, 2013, from http://www.aapb.org/i4a/pages/index.cfm?pageid=3463

Baños, R. M., Guillen, V., Quero, S., García-Palacios, A., Alcaniz, M., & 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.

Bassotti, G., & 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.

Benson, H. (1975). *The Relaxation Response*. New York, NY, USA: HarperCollins.

Bersak, D., McDarby, G., Augenblick, N., McDarby, P., McDonnell, D., McDonald, B., & Karkun, R. (2001). Intelligent biofeedback using an immersive competitive environment (pp. 1-6). In: *Proceedings of UbiComp 2001 Workshop on Ubiquitous Gaming*.

Bouchard, S., Bernier, F., Boivin, É., Morin, B., & Robillard, G. (2012). Using biofeedback while immersed in a stressful videogame increases the effectiveness of stress management skills in soldiers. *PLoS ONE*, 7(4), e36169.

Boucsein, W. (2006). *Electrodermal Activity* (2nd ed.). New York, NY, USA: Springer.

Buckelew, S. P., Conway, R., Parker, J., Deuser, W. E., Read, J., Witty, T. E., et al. (2005). Biofeedback/relaxation training and exercise interventions for fibromyalgia: a prospective trial. *Arthritis & Rheumatism*, 11(3), 196-209.

Carr, A. (2006). *Handbook of Child and Adolescent Clinical Psychology: A Contextual Approach* (2nd ed.). London, UK: Routledge.

Chandler, C., Bodenhamer-Davis, E., Holden, J.M., Evenson, T., & Bratton, S. (2001). Enhancing personal wellness in counselor trainees using biofeedback: an exploratory study. *Applied Psychophysiology and Biofeedback*, 26(1), 1-7.

Coolican, H. (2009). *Research Methods and Statistics in Psychology*. London, UK: Hodder.

Critchley, H. D., Melmed, R. N., Featherstone, E., Mathias, C. J., & 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., & 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., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, 7(2), 189-195.

Dekker, A., & Champion, E. (2007). Please biofeed the zombies: enhancing the gameplay and display of a horror game using biofeedback (pp. 1-9). In: DiGRA 2007.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17(2), 124-129.

Enck, P. (1993). Biofeedback training in disordered defecation. *Digestive Diseases and Sciences*, 38(11), 1953-1960.

Fentress, D. W., Masek, B. J., Mehegan, J. E., & Benson, H. (1986). Biofeedback and relaxation-response training in the treatment of pediatric migraine. *Developmental Medicine & Child Neurology*, 28, 139-146.

Fleureau, J., Guillotel, P., & Huynh-Thu, Q. (2012). Physiological-based affect event detector for entertainment video applications. *IEEE Transactions on Affective Computing*, 3(3), 379-385.

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), article 0001a.

Friedman, D., Suji, K., & Slater, M. (2007). SuperDreamCity: an immersive virtual reality experience that responds to electrodermal activity. In Paiva, A., Prada, R., & Picard, R.W. (Eds.): *Affective Computing and Intelligent Interaction, LNCS 4738*, 570-581. New York, NY, USA: Springer.

Golz, J., & Kaernbach, C. (2013) *Ledalab*. www.ledalab.de (last accessed on 18 September 2013).

Gorini, A., & 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(25), 25-33.

Healey, J. A., & Picard, R. W. (1998). StartleCam: a cybernetic wearable camera (pp. 1-8). In: Proceedings of the Second International Symposium on Wearable Computers (ISWC 1998). Washington, DC, USA: IEEE Computer Society.

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2), 156-166.

Hunyor, S. N., Henderson, R. J., Lal, S. K. L., Carter, N. L., Kobler, H., Jones, M., et al. (1997). Placebo-controlled biofeedback blood pressure effect in hypertensive humans. *Hypertension*, 29, 1225-1231.

Jensen, M. P., Barber, J., Romano, J. M., Hanley, M. A., Raichle, K. A., Molton, I. R., et al. (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.

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., & 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., & 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.

Lisetti, C. L., & 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., & 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.

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., & 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.

Mandryk, R. L. (2005). *Modeling user emotion in interactive play environments: a fuzzy physiological approach.* Doctoral dissertation. Simon Fraser University.

Mandryk, R. L., & 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.

Nasoz, F., Lisetti, C. L., & Vasilakos, A. V. (2010). Affectively intelligent and adaptive car interfaces. *Information Sciences*, 180(20), 3817-3836.

Pallavicini, F., Cipresso, P., Raspelli, S., Grassi, A., Serino, S., Vigna, C., Triberti, S., Villamira, M., Gaggioli, A., & Riva, G. (2013). Is virtual reality always an effective stressor for exposure treatments? Some insights from a controlled trial. *BMC Psychiatry*, 13(52).

Picard, R. W. (1997). *Affective computing*. Cambridge, MA, USA: The MIT Press.

Picard, R. W., & Goodwin, M. S. (2008). Developing innovative technology for future personalized autism research and treatment. *Autism Advocate*, 50(1), 32-39.

Pistoia, F., Sacco, S., & Carolei, A. (2013). Behavioral therapy for chronic migraine. *Current Pain and Headache Reports*, 17(1), 304-311.

Rani, P., Sarkar, N., Smith, C. A., & Adams, J. A. (2007). Anxiety-based affective communication for implicit human-machine interaction. *Advanced Engineering Informatics*, 21, 323-334.

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., & 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., & 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., & Alcañiz, M. (2007). Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavor* 10(1), 45-56.

Rupert, P. A., & 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.

Scheirer, J., Fernandez, R., Klein, J., & Picard, R. W. (2002). Frustrating the user on purpose: a step toward building an affective computer. *Interacting With Computers*, 14(2), 93-118.

Shusterman, V., & 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., et al. (2006). Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence*, 15(5), 553-569.

Tsai, P. S., Chang, N. C., Chang, W. Y., Lee, P. H., & 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.

Vasudeva, S., Claggett, A. L., Tietjen, G. E., & 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.

Villani, D., & 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.

Villani, D., Riva, F., & Riva, G. (2007). New technologies for relaxation: the role of presence. *International Journal of Stress Management*, 14(3), 260-274.

Wu, D., Courtney, C. G., Lance, B. J., Narayanan, S. S., Dawson, M. E., Oie, K. S., & 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.

Zhai, J., & 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., & Mark, R. (2003). An open-source algorithm to detect onset of arterial blood pressure pulses. *Computers in Cardiology*, 30, 259-262.