# Chapter 4
# Games User Research and Physiological Game Evaluation

**Lennart E. Nacke**

**Abstract** This chapter introduces physiological measures for game evaluation in the context of games user research (GUR). GUR consists of more than playtesting game; it comprises a collection of methods that allow designers to bring their creations closer to the initial vision of the player experience. With the prices of physiological sensors falling, and the advancement of research in this area, physiological evaluation will soon become a standard tool in GUR and game evaluation. Since mixed-method approaches are of increasingly prominent value, this chapter describes core GUR methods with a special focus on physiological evaluation, keeping in mind both benefits and limitations of the approach in academic and industrial applications.

## 4.1 Introduction

From the academic domains of human-computer interaction, human factors, and social psychology, robust and scientific user-testing approaches have been adopted in a game industry field called games user research (GUR) to ensure optimal quality of the user experience (UX) in games and virtual entertainment products. In the games industry, the domains of quality assurance (QA) and game testing are focusing on finding technical errors in the game code (i.e., bugs) and ensuring smooth execution of the game on a technical level. By contrast, GUR is concerned with evaluation through the observation and analysis of players. A game designer communicates their thoughts to the player using the game. In turn, the user researcher applies methods that are inspired by psychology and user-centered design to evaluate the player. The communication channel is mutual, and allows for interpretation of player reactions (and often questions) from the user researcher to inform the designer what features of the game could be improved. Figure 4.1 illustrates the flow of information resulting from communication channels at all stages of the game

L. E. Nacke (✉)
Faculty of Business and Information Technology, University of Ontario Institute of Technology, Ontario, Canada
e-mail: lennart.nacke@acm.org; web: www.hcigames.com



R. Bernhaupt (ed.), *Game User Experience Evaluation,*
Human-Computer Interaction Series, DOI 10.1007/978-3-319-15985-0_4

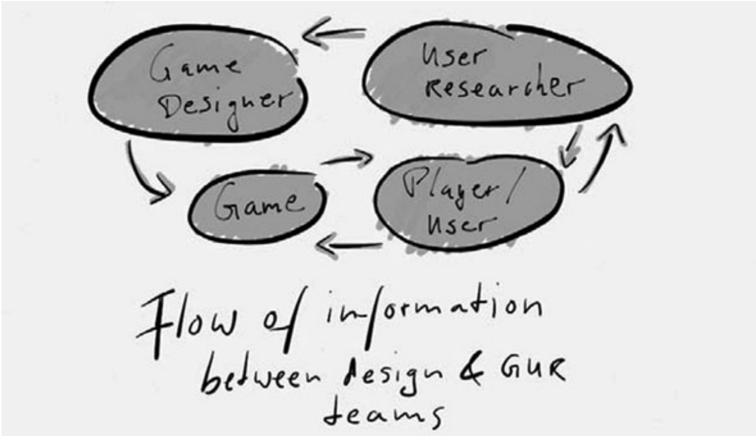


**Fig. 4.1** Flow of information between game design and user research teams

creation and evaluation process. This allows for communication between player, designer, user researcher and the game (the designer's communication tool).

Notably, the focus of GUR in industry lies on experience analysis and understanding player interaction, with the objective of not simply testing the player, but improving the game's design. Seasoned GUR professionals describe their job as well done when they can provide game designers with an (often frustrating) moment of insight about how their designs are being interpreted by players. In this way, GUR can take inspiration from the scientific method as an approach of evaluating design hypotheses that are created during each development cycle in a game. Ultimately, GUR aims to allow designers to create more compelling gameplay experiences by identifying weaknesses in the design and structure of a game prototype. Some of the most advanced methods for GUR, such as physiological evaluation and methods using expensive recording and analysis equipment, are currently used primarily in a laboratory setting (Nacke 2013). They are likely to become more popular within the next years, because sensors are becoming more affordable as technology advances. In the context of these advanced GUR methods, this chapter will focus on the physiological evaluation of players, which takes inspiration from psychophysiological research.

Psychophysiology is a research area where body signals, or so-called "physiological responses", are measured to understand their underlying psychological processes, and the connections between those processes. I will refer to this as physiological evaluation in this chapter. Essentially, we are studying the signals that our body produces (the physiological part) to get an idea of what our mind was doing at that point (the psychological part). This makes physiological evaluation a useful method for evaluating arousal/excitement, emotion, or mental workload in games, although valid academic studies or experiments require much caution, as I will explain later. Most of our body signals or responses are spontaneous, which means that they are difficult to fake, making physiological measures highly objective. This lends the technique a comparatively low level of bias when analyzing a player's reactions to gameplay. They are also continuously recorded without interrupting a player's gameplay session. Physiological metrics consist of vast amounts of data that become meaningful only in the correct context (Mandryk 2008; Nacke 2009). For example, as game designers, we want to create meaningful decisions that involve some form of tradeoff in terms of game resources (e.g., resource trades, weighing risk against reward, and choosing an appropriate action) (Brathwaite and Schreiber 2008). Emotional decisions are one of the primary fun factors in playing games, because they add value to the outcomes of our decisions. In these decision situations, physiological metrics allow for an objective way to assess a player's emotional and cognitive response (in addition to the behavioral response, which can be observed or inferred from gameplay interaction logs). In the case of our example, this refers to whether the designers have succeeded in causing an emotional response in the player with the decision options that they have provided. For example, an experiment could compare whether it is more interesting (emotionally) to buy a powerful sword that does the same amount of damage and has the same price as a comparably effective attack scroll. However, one has to keep in mind that quantitative physiological data has to be interpreted to make correct design suggestions, which leaves room for interpretation bias on the part of the researcher.

If we were to roughly distinguish physiological games user research in industry and academia, we could craft a series of generalizations separating the two approaches. In the *game industry*, games user researchers are using psychophysiological methods to evaluate, for example, emotional and cognitive effects of game design (Ambinder 2011). In *academia*, human-computer interaction and psychology researchers are often using psychophysiological methods to study concepts of user interaction or effects of game content on player aggression (Carnagey et al. 2007). Academics emphasize using controlled experimental conditions and participant samples, place importance on statistical power, and carefully remove order effects by counterbalancing the design. Especially for experimental research psychologists, the game is a means to an end. It provides the virtual environment or engaging task that is used to study player behavior. From an experimental psychology viewpoint, one main difference between academic and professional approaches is that psychophysiological research is more concerned with experimental validity, underlying scientific questions, and controlled conditions, while GUR professionals are more interested in quick insights regarding game design or interface design that are immediately actionable. A new breed of GUR researchers in HCI is trying to combine the best of both worlds: Providing actionable and relevant results without sacrificing experimental validity. For example, if a think-aloud protocol is applied when recording physiological metrics, a researcher risks influencing heart rate and respiration considerably. External physical activity can create a physiological response in similar ways as internal physiological activity; this can lead to environmental effects influencing results.

It has to be emphasized here that physiological data is volatile, variable, and difficult to interpret without a high level of experimental control. When interpreting physiological metrics, it is also important to understand the relationship between psychological effect and physiological response. I am using the term physiological response to refer to different physiological response signals (e.g., brainwaves, muscle activity, or skin conductance). The most common relationship in physiological evaluation is the many-to-one relationship, where one physiological response may be associated with many psychological effects (Cacioppo et al. 2007). Therefore, it is important to keep in mind that a direct mapping of an emotional state to a psychological effect is not possible, and physiological responses must be understood as elements of sets with fuzzy boundaries. When we measure physiological signals, we are measuring essentially the operation and activity of muscles, nerve cells, and glands (Stern et al. 2001).

## 4.2 Games User Research Methods

Before we focus on physiological evaluation, let us have a look at the currently common non-physiological GUR methods: behavioral observation, think-aloud protocol, interviews, heuristic evaluation, focus groups, surveys and questionnaires, and game metrics. While these traditional methods are still the industry standard, some newer methods are being developed currently that combine traditional approaches of user testing (Mirza-Babaei et al. 2013).

### *4.2.1 Behavioral Observation*

This technique can be as simple as looking over the shoulders of individuals when they are playing a game, or as complex as high-definition video recording of players from different angles in a natural gaming environment, such as a living room. Because of its relative simplicity and the actionable results that it produces, behavioral observation is often regarded as the most valuable GUR technique. Direct observation is especially valuable to the designers of a game. Seeing how players deal with the game's challenges, or where they get stuck or frustrated in the game level, can lead to profound insights regarding the gameplay experience for game designers. However, some inference is required on the designer's part, since observation only tells them *what* is happening when players play their game; it does not answer *why* something is happening to a player, or how the player felt when it was happening. Another valuable observation is players' body language and facial expressions while they play the game. It is important to note that during observation, the researcher should only take notes of the things that are being observed, without including any premature inferences. This is why there should always be a GUR experimenter[1] present when designers are observing players. It is not until discussion with the GUR professional after the observation that points of improvement

[1] Note that we use the term "experimenter" here to refer to a trained GUR professional or physiological researcher in charge of running the experiment.

**Table 4.1** An example protocol for behavioral observation

| | |
|---|---|
| 1 | *Design the experimental session*. Decide what parts of the game and the player you want to observe. This has to be a small and modifiable part of your game. |
| 2 | *Write an experimental script*. Following a script in general allows a formalization of this test approach (and is necessary in many other methods as well). The script needs to explain where the starting point for the player is, since a session needs clearly defined start and end points. Make sure you have software or a notebook ready for recording your observations. |
| 3 | *Think about what behaviors could be expected*. It helps to label reoccurring behaviors, and makes conversation with other observers easier. In behavioral observation this catalogue of common nomenclature is called an *Ethogram* (van Abeelen 1964). |
| 4 | *Test the script and setting in a pilot run-through*. Make sure to get feedback from a team member regarding the clarity of your instructions. |
| 5 | *Define your target demographic and start recruiting*. It helps to have a narrowly defined target group of players for your test. Once the limiting parameters of this demographic have been defined, it is time to recruit people. |
| 6 | *Run the session*. Make sure to sit outside the field of view of your players and record the gaming session. Make detailed notes (even if you are using recording software, since it is always good to have a backup of observations), not only about what the players are doing, but at what time of your session this behavior occurred. Be sure to only write the observation down and not the interpretation (e.g., "the player is sighing and frowning" instead of "the player is frustrated"). It is also important to log all the performance measures the game gives you (e.g., time taken for a level, high score). Especially if you are not using metrics logging, this information is invaluable for your analysis. |
| 7 | *End of session protocol*. You should have a procedure in place to thank players for their time, and give them some more information about the session if required. However, this information must be neutral in tone, without any value judgments about a player's performance. |
| 8 | *Schedule for break time*. In case there are multiple sessions scheduled in your day, make sure to allow for breaks in between your sessions to reset everything and ensure a relaxed mindset and comfortable environment for the next participant. |

can be clearly identified. It is best to have one or two observers focusing on a single player, so that neither player nor observer is distracted during the process. During observation, communication and any interruption of gameplay should be avoided (if necessary, only neutral communication should occur between experimenter and player). This requires skill and patience on the observer's part, for example, not to influence the player when they get stuck (Table 4.1).

Becoming a good observer takes time and practice, so even though the observation method is easy to use, there is some training time involved. Like any subjective interpretation method, there is some bias arising from the experimenter's interpretation of observational data (which some might argue is stronger when interpreting observational vs. numeric data). Finally, most of the time, there are actionable results to be gathered from behavioral observation, especially when a game designer is present to discuss the next iteration steps together with the user researcher.

While behavioral observation is most often used in an industry setting, one could use the same method to investigate an academic hypothesis. A research question would likely be targeted to explore how participants react to the events that the game presents. An analysis would need to investigate the occurrence and quality of the behavioral cues observed.

### *4.2.2 Think-Aloud Protocol*

The think-aloud protocol was developed in usability testing for products (Lewis 1982) and later made its way from interaction design to GUR. It could be seen as a natural expansion of behavioral observation, because it introduces player narrative to the observations. Players are asked to talk about what they are thinking as they play through the game. This voice over is often (and ideally) recorded, so that later in the analysis, the researcher can draw conclusions not only based on behavior, but also on player descriptions. Similar to behavioral observation, it is important to let the player speak and not interrupt the train of thought. It is also important that the speaking is natural, so that the players know that they do not have to talk and think about what to say. The analysis of the narrative has to happen after the session.

This method requires some skill from the player (for unskilled think-aloud participants, a moderator might need to be present to trigger player reporting, which could influence the natural flow of information). The better the player is at commenting their own behavior in the game, the more insights this method is likely to yield for the GUR professional. Talking about what one is doing is not natural for many people, and will likely require some practice, so that it is useful to make a note of which players have been particularly good at commenting, and inviting them again when another game is being tested. A skilled moderator can encourage players to speak freely, but adds another person to the testing session (since the GUR professional should only observe and make notes of the behavior), which adds some overhead on the planning side.

### *4.2.3 Interviews*

Interviews are a common method of qualitative subjective inquiry, providing direct insight into the player experience. Much of the quality of the data collected during an interview session depends on the skill of the interviewer. Interviews allow a greater degree of depth when analyzing player opinions, emotions, and reactions. However, data collection and analysis can be time-consuming.

Interviews are a classic GUR method allowing for rich data collection if done correctly, providing the opportunity to increase the specificity and accuracy of data with follow-up questions. However, the analysis procedure is time-consuming, and might not produce patterns in responses that can be categorized into common themes. These common themes can help quantify the responses from multiple individuals to produce more reliable conclusions from interviews. Since human memory is limited, sometimes a gameplay video (recorded during the play session) is used

**Table 4.2** An example of an interview session

| | |
|---|---|
| 1 | *Set up a calm environment*. To get the most out of interview data, it is important that you and the person you are interviewing can focus on the questions at hand. Therefore, try to find a comfortable and distraction-free setting to conduct the interview. |
| 2 | *Get recording aid*. While it is possible to conduct the interview with just you and your interviewee while you are taking notes, you might lose interesting data, since it is impossible to capture everything that is being said in your notes. It is highly recommended to use at least an audio recorder for an interview. This way you can focus on the interview, and do not have to worry about what to write down. Like any other discussion with players, it is important that you make sure it is understood that the game is being analyzed and not the players. |
| 3 | *Prepare your questions and interview script*. Like any other experimental GUR method, a script will help you executing your interviews more smoothly. At the very least, you should have a set of questions prepared for the interview session. Writing non-biased questions is another skill that takes some time to develop. Ideally, the questions should aim at the how and why of player actions to get the most insight out of an interview. For simple yes or no questions, a questionnaire might be the more appropriate method. |
| 4 | *Conduct the interview*. Start with an easy question to establish a good atmosphere, and then drill deeper regarding certain behaviors that you are interested in. It is important to acknowledge that you understand and listen to what is being said, so that the interviewee feels good about talking to you. At the very end of an interview, it is good to allow for some general questions about the process, or to ask the players to add anything they feel strongly about or that they would really like to talk about. |
| 5 | *Transcribe your interview*. This will be time consuming, as you have to revisit all the audio that you have recorded and find patterns in the answers of different interviewees. Software can help a lot in this process, but it is ultimately up to the researcher to find common themes. |

as a recall aid to trigger the memory of a player's experience (Table 4.2). Finally, interviews are a subjective GUR method, since they allow for interpretation and answer biases.

### *4.2.4 Questionnaires*

Questionnaires or surveys are a common GUR method, because they allow collecting large volumes of self-report data simultaneously from many different players. Surveys are usually used to get insights into value judgments about gameplay moments. Again, this book provides two excellent chapters about survey construction: Cavillo-Gamez et al.'s Chap. 3 about the core elements of the gaming experience, and Takatalo et al.'s Chap. 5 about presence, involvement, and flow.

A questionnaire can be delivered to players directly after a gameplay event (e.g., making a choice in the game) or after a gameplay session, so that experience is still present in memory. This is arguably less biased than, for example, recalling an experience during an interview session post-gameplay. Often, gameplay questionnaires feature a Likert-scale (Likert 1932) rating type of a gameplay interaction item. Surveys can also facilitate the understanding of a player demographic or psychological type (Eysenck et al. 1985; Carver and White 1994; Buss and Perry 1992; John and Srivastava 1999) and can be delivered before a play session to assess a gamer type (Bartle 1996; Nacke et al. 2014). Some of the more popular surveys used within GUR evaluations are focused on immersion (Jennett et al. 2008), emotion (Lang 1995), and game engagement (Brockmyer et al. 2009). Surveys allow the construction of subjective metrics, which can be valuable for improving insights into biometric and game metric data. While questionnaires provide a quick way to obtain quantitative insights into player feelings and attitudes, they lack the depth of an interview or the objectivity of metrical measures. They also work most reliably when a large number of people are available for testing.

### *4.2.5 Focus Groups*

Focus groups are another method of qualitative inquiry, where a group of players is gathered to talk about their opinions, beliefs, and attitudes towards the game (Poels et al. 2007). The group has complete freedom to talk about their likes and dislikes, but usually a moderator is present to facilitate the discussion and lead the group toward a topic of interest. Again, I will not go into depth about focus group methodology here, since Poels et al. (2010) describe a focus group study about qualitative insights into postgame experiences, which is an excellent overview of the methodology.

Since focus groups are an easy way to gather opinions and feedback on ideas, they can be used early on in development, even after a game prototype session or an initial design presentation. While focus groups allow feedback from crowds rather than individuals (somewhat seeming like an extension of the interview technique), they have the disadvantage that once an opinion is present in the crowd, others may pick up on this opinion and the results may become biased. Imagine, for example, that only a couple of strong voices dominate a focus group. This can lead to group pressure, preventing other individuals in the group from voicing their own opinions about their gameplay experience. Sometimes, focus groups skew toward the discussion of a particular solution rather than focusing on issues regarding player experience. In the end, this method might not be as useful as other GUR methods, due to the limited actionability of pure opinions, as opposed to the more complete data offered by methods involving observation or standardized metrics. A GUR professional is usually more interested in what players are actually doing when playing the game (not so much in what they think they want to do in the game).

### *4.2.6 Heuristic Evaluation*

Heuristic evaluation is a method stemming from usability research (Nielsen and Molich 1990). In this context, the evaluation consists of judging how an interface complies with recognized usability principles, which are called “heuristics”. It is known as a discount usability method, because it is cheap to conduct and can yield significant actionable results for a game. The method includes GUR experts playing a game and evaluating it based on a set of criteria. So, in a way, a heuristic evaluation can be likened to a tightly structured game review. After the expert has played the game, they give feedback on whether the game fits a certain playability guideline, or "heuristic", and what problems might arise from non-compliance. Several different heuristics have been proposed by researchers (Korhonen and Koivisto 2006; Korhonen and Koivisto 2007; Jegers 2008; Koeffel et al. 2010; Pinelle et al. 2008a) and practitioners (Desurvire et al. 2004; Desurvire and Wiberg 2008). We will not get into details about game heuristics here, because Chaps. 8 and 9, Part III of this book by Desurvire and Wiberg as well as Hochleitner et al. already cover much in-depth information about the topic.

In general, a heuristic might be more powerful if it fits a certain game platform (e.g., mobile) or genre. While there is no commonly agreed upon set of heuristics, certain overlaps exist, such as the game having a clear goal and an understandable control scheme. Other heuristics might concern the difficulty level, fair outcomes, learning curves, and repetitiveness. These guidelines may seem like common sense, but often they are not followed, and using written rules for the structural evaluation of games makes the evaluation process easier. It is important to note some useful modifications to traditional heuristic approaches that we have seen in recent years, especially the genre weightings proposed by Pinelle et al. (2008b) and the critic proofing approach developed by Livingston et al. (2010), which produces a list of heuristic violations (taking into account a problem's frequency, impact, persistence, the heuristic it violates, and a game's genre).

A definitive advantage of heuristic evaluations is that only a small number of experts are needed for evaluating the game. However, the selection of these experts can already pose a problem, since each person needs to have expertise relevant to the game at hand, and should use heuristics that are directly applicable to the game being evaluated. Another issue is the subjective bias of the evaluators, which has been shown to cause low agreement in the issues and causes detected by evaluators (White et al. 2011). In some cases, an expert might still miss a problem that is relevant for a novice player. Ideally, this method is combined with the observation of novice players, to extend the feedback gathered from the experts.

### *4.2.7 Game Metrics*

Game analytics and metrics are undeniably a recent trend within the GUR community (Seif El-Nasr et al. 2013). Chapter 7 by Drachen in this book focuses on game metrics or more specifically on a part of game metrics that is also referred to as game telemetry (because data are often collected after delivery of the game and not always during testing). The term "game metrics" generally refers to the process of logging player interactions, positions in the game world, camera angles, and all data that relates to the gameplay interaction process in the game. To do this, a programmer has to define appropriate hooks in the game engine that allow the logging of all this data. A huge advantage of metrics is the large amount of data being collected. This can also prove detrimental, as this data has to be analyzed quickly and accurately, which often means that a GUR professional has to use visualization software to make sense of such vast collections of data. The potential of this GUR methodology is promising, because it can measure key gameplay events (e.g., player deaths) and their surrounding circumstances in detail. By integrating these log files with synchronized physiological sensor data, we can craft a more complete picture of player experience.

## 4.3 Physiological Game Evaluation

Quantitative data is not only available from in-game logging, but can also be acquired directly from players. For a deeper understanding of the complex physiological processes involved in decision-making in games or emotional processing of game events, we can use physiological measures to evaluate user experience (UX) aspects of games. It has to be understood—before we start—that physiological measures do not measure UX itself, but tap into parts of UX that include measurable physiological components (Kivikangas et al. 2010). Various physiological measures are not covered in this chapter, such as respiratory sensors, eye trackers, temperature sensors, and several brain imaging techniques. A part of this introduction to physiological measures has previously appeared in Nacke (2013).

### *4.3.1 Introduction*

Physiological signals are small energy measures gathered from the surface of our body. We need to understand how our body operates on a neurobiological level to better understand the meaning of these signals. On a macro level, the human nervous system controls the operations of the body. This is split into two parts: the central nervous system (CNS) and the peripheral nervous system (PNS). The CNS manages any information received from the whole body, coordinating activity accordingly. The skull and spine bones protect the CNS, which also makes it difficult to access outside of the body. The PNS includes all nerve cells outside of the CNS, and it connects the CNS to the rest of our body.

The PNS transmits most of our physical sensations. This exposes it to measurements on our skin. Consequently, the skin surface is the location from where physiological sensors commonly get their signal. The PNS is further divided into the somatic and autonomic nervous system. The somatic nervous system regulates body activity that we have under conscious control, such as deliberate muscle activity. The autonomic nervous system (ANS) controls our unconscious, visceral responses. We have two opposing players in the ANS, the sympathetic nervous system (which triggers fight or flight reactions in emergency situations) and the parasympathetic nervous system (which controls our relaxation, resting, and digestion). It is important to keep those two systems in mind when we measure emotions with physiological sensors.

Physiological and mental processes contribute equally to the formation of emotions (LeDoux 1998). We associate emotions with feelings, behaviours, and thoughts. Recent research supports a strong connection between affective and cognitive processes as a foundation for emotions (Pessoa 2008; Damasio 1994). A popular emotion model currently used in psychophysiological research is Russell's circumplex model (Russell 1980) that suggests two emotional dimensions (see also (Larsen et al. 2001; Larsen and Diener 1992)), pleasant/unpleasant emotion (or valence) and high/low stimulation (or arousal, excitement). When we are trying to measure stimulation or arousal, the PNS is more reliable than when we are trying to measure emotional valence. However, by using measurement techniques that detect muscular movement in the face (e.g., an activation of the cheek or the brow muscle), we can measure emotion based on facial expressions. According to the psychological literature, negative emotion manifests in a frowning facial expression and positive emotion is displayed as a smiling face (Cacioppo et al. 1990). As games user researchers, we are interested in feelings (or how players interpret their emotions in context), and player experience. Therefore, currently we cannot solely rely on physiological metrics for player testing; we have to accompany these methods with other GUR techniques (e.g., the techniques discussed at the beginning of this chapter) to fully understand player experience. However, the basic tenet of physiological experimentation still holds true: we measure the physiological response (in addition to other subject responses) while manipulating a behavioural factor, often an element of gameplay. To have a better understanding what physiological metrics are capable of, we are going to discuss the most common ones in the following sections. Since the experimental procedure for physiological evaluation in a laboratory setting can be similar, I am giving a sample overview of an experimental protocol here (see Table 4.3).

### *4.3.2 Electromyography (EMG)*

EMG measures whether or not our muscles are active. Therefore, an EMG electrode attached to the surface above a muscle is able to sense even the slightest activation of this muscle (Bradley et al. 2001; Lang 1995). Whenever we flex a muscle on our body, this produces a difference in electrical activity or isometric tension which is measurable by EMG. EMG is all about measuring PNS activation, in contrast to other techniques focused on CNS activation (such as EEG, which measures brainwave activity). Since most muscles can be directly controlled, EMG is a measure of high interest for interacting with computers in a more natural way (Nacke et al. 2011b). The most prevalent use of this technique in evaluating games is through facial EMG (Fridlund and Cacioppo 1986), which measures the activation of specific facial muscles responsible for displaying our positive or negative reactions to an emotional moment in a game (Hazlett 2006). In particular, physiological game research has focused on using brow muscles (corrugator supercilii) to indicate negative emotion and cheek muscles (zygomaticus major) to indicate positive emotion (Mandryk et al. 2006) or even fun and flow in a game (Nacke and Lindley 2008).

**Table 4.3** An sample experimental protocol for physiological evaluation

| | |
|---|---|
| 1 | *Check lab inventory*. It is a good idea to have a lab inventory set up for physiological experimentation, since these experiments usually require disposable items to be used in every session (e.g., pre-gelled electrodes, surgical tape, electrode paste). The inventory should always be up to date (and checked) at least a week before experiments start, so that missing items can be ordered in. Most physiological systems operate with batteries, so it has to be checked that the physiological recording unit is fully charged before the experiment. |
| 2 | *Prepare informed consent forms*. These should be required for any kind of experimentation, but are most valuable for physiological experimentation, to inform the participants that they will not be harmed by the measures and that they can opt out at any time. These forms often fulfill the role of both ethical documentation and participant education (no mind-reading or graphical access to thoughts). Participants need to be informed before participation that they should not consume substances that influence physiological responses (e.g., coffee, energy drinks, candy). |
| 3 | *Prepare and attach the electrodes*. The differences between physiological recording systems will require different preparations. In general, it is good to keep in mind that pre-gelled and dry electrodes are generally much easier to handle (and will decrease the total time needed for your experiment) than electrodes that require paste. If EEG is being used, a cap will facilitate the proper alignment of electrodes. It is good practice to mark the channel names (usually numbers), so that it is easy to identify electrode locations for multichannel systems. (For example, which EMG electrodes were measuring what muscle site.) |
| 4 | *Check for noise and start recording*. Often it is quite easy to identify the noise in physiological signals, as they will look jittery in the recording software or exhibit unnatural spikes. Many factors can influence this; some of the common culprits are cable issues (broken or bent), electrolyte bridges (too much gel), and lack of skin contact (not enough gel). It is good practice to check for a couple of potential causes and repeat the attachment procedure until the error disappears. At the start of the recording, participants should sit in an eyes-open and calm position, so that a baseline can be recorded (which is necessary for the later normalization of data). Depending on the software used, recording files may differ, but it is good practice to have a numbering scheme that allows sorting and comparison of the data that is being recorded. |
| 5 | *Run experiment, follow up with additional measures*. After recording is set in place, the regular experiment can be run. Ideally, the participant's behavior should be observed, and markers put in place when the observer notices too much movement from the participant (which can lead to movement artifacts in the physiological recording). Surveys can provide a good addition (just like any other GUR method discussed in this chapter) to the physiological data, and are often administered after the session. |
| 6 | *Debrief and back up data*. After the experiment is over, participants are generally thanked and debriefed about the study. Given that physiological files are generally large and will grow over time as more data is being collected, it is a good idea to back up the data after the experiment is over. |

For longer term evaluation (say over a few minutes of gameplay), the eye muscle (orbicularis oculi) has also proven helpful in registering high arousal pleasant emotions (Ravaja et al. 2008).

Similar to EEG, EMG uses silver-silver chloride electrodes (see Fig. 4.4) because they have only a small measurement error, little drift potential, and minimal polarization. EMG electrodes are applied to the surface of the skin and will also need a reference (if part of a larger system, this reference can be on head or close to the actual electrodes, which are often arranged in a triangular patch). The risk of facial EMG is the presence of considerable muscle activity that you possibly do not want to measure (e.g., experimental participants cannot chew gum, laugh, or talk during facial EMG, because this will introduce large artifacts into your data). Since muscular signals are again amplified from microvolts, careful signal processing has to be done on EMG data before it is interpreted (e.g., a log normalization or a Butterworth low-pass filter).

Emotions are usually interpreted on a two-dimensional model (Russell 1980). We find that by measuring facial muscles, we are able to get an idea of pleasant or unpleasant emotions. This is called valence assessment, as we are able to say whether an emotion was evaluated by a player as pleasant or unpleasant. The other dimension of emotional measurement, arousal, serves as an indicator of player excitement. While facial recognition software or direct observation would also allow the analysis of facial expressions and therefore the mapping of emotions, the software or the observer often miss less salient expressions, which are picked up by physiological measures. The analysis of EMG signals is pretty straightforward, as usually after application of a notch filter, we are already able to compare the signals. However, these measures require the attachment of electrodes to the player's face, which make this method somewhat intrusive, although often the electrodes and cables can be easily taped to the player's head to remove discomfort and reduce movement artifacts. Since just by feeling the electrode on their face, players might be feeling the need to elicit more pronounced muscle movements, this might lead to unnatural signals, which could make data interpretation more difficult.

### *4.3.3 Electrodermal Activity (EDA) and Galvanic Skin Response (GSR)*

When measuring the level of skin conductance over time, we refer to this as measuring the EDA of the skin, but when measuring the direct response to a stimulus, we refer to this as galvanic skin response (Boucsein 1992). In any case, EDA measures changes in the passive electrical conductivity of the skin, relating to increases or decreases in sweat gland activity.

Most of us have seen EDA measures in movies, branded as lie detector tests. EDA measures are attached to the fingers, palms or toes because the sweat glands in those body areas are more likely to react to changes in the PNS (sympathetic vs. parasympathetic activity). Since we are measuring the differences in conductivity, we only need two electrodes, which make EDA a very easy physiological measure to prepare and apply. EDA electrodes are prone to movement artifacts just like other psychophysiological measures, but because of their location (hand or feet), special care has to be taken in the preparation of steering controls for the game. If a regular game controller is used, and the EDA sensor is applied on the palm of the hand, movement artifacts are likely to occur, so using the feet or fingers of the non-dominant hand would be a better location (or the side of the palm of the hand). EDA is also very easy to interpret, since it almost has a one-to-one relationship with physical arousal. Similar to the EMG analysis, EDA values need to be normalized to allow for a meaningful statistical comparison between participants in your study. EDA can be normalized by calculating each sample as a percentage of its entire span, using the min and max values over all samples for one participant (Mandryk 2008; Lykken and Venables 1971).

Another benefit of this measure is the inexpensive hardware that usually comes at a fraction of the cost of a research-grade EEG setup. Many modern EDA systems use dry electrodes, and some EDA setups even allow easy attachment of the electrodes to the little and ring fingers with a Velcro strap.

A benefit of EDA is that analysis can be done in a macro (over larger chunks of playtime) or micro fashion (related to events). When analyzing the response to a direct event, one needs to take into account that EDA is a relatively noisy signal that also has some latency in response to a stimulus (often around 5 seconds, see Fig. 4.5). After a galvanic skin response is registered, there is also a decay or recovery time during which no further event responses will be registered (or the responses are registered together). While it is pretty clear that EDA indicates physical arousal, there is still some interpretation effort required as to the source of the initial stimulation (e.g., caused by a game event, or influenced by non-game environmental factors). This is why planning and controlling physiological experiments is very important. Since EDA also has large individual variations in baseline and responsiveness of the signal, it is very important to clean up the raw data with baseline corrections and other filtering techniques.

### *4.3.4 Cardiovascular Measures*

There are many cardiovascular measures available for physiological evaluation, and all of them relate to the heart rhythm, its changes, and how this influences one's physiological state. The most common measures are electrocardiography (ECG), heart rate (HR), Interbeat Interval (IBI), heart rate variability (HRV), blood volume pulse (BVP), and blood pressure (BP). However, blood pressure is not a real-time measure, and, while useful in medical contexts, has not shown relevance in games, user research. For all other relevant measures, physiological electrodes are required to collect the necessary data.

ECG measures the electrical activity caused by the heart pumping blood and is measured usually with three electrodes or leads, which are positive, negative, and neutral and are usually applied to the upper body.

Heart rate is understood as the number of heart beats per time unit (usually measured in beats per minute). The IBI, or time between heart beats, is also of interest. If IBI decreases, HR obviously increases, and this has been tied to increased information processing and emotional arousal. So, IBI and HR are two related measures. However, HR variability is a more complicated measure (with a complex analysis procedure) as it relies on the spectral analysis of sinus arrhythmia. In plain English, we are looking at frequency changes and differences in the IBI over time. In general, we need to keep in mind that cardiovascular measures are intrusive to measure accurately, and that they are affected by many things, such as physical activity.

### *4.3.5 Electroencephalography (EEG)*

It is fascinating to see a computer plot out the frequencies and amplitudes of one's brainwaves in real time. While some relationships between EEG activity and behavior have been established, there is still much left to be researched, especially with complex stimuli like video games. Participants unfamiliar with EEG might assume that an experimental researcher is able to find out exactly what they are thinking, or obtain graphical representations of their thoughts. Recent research has investigated the latter (i.e., reconstructing visuals from brain activity using magnetic resonance imaging [MRI]) (Nishimoto et al. 2011). However, the reality of EEG measures is a little more abstract. EEG, which uses electromagnetic measurement, can be considered less invasive and easier to use compared to other CNS response analyses, such as functional MRI, which analyses brain function based on hemodynamic response[2], or positron emission tomography (PET) scans, which use metabolic response to display brain function. A major advantage of EEG for brain activity measurement over these other techniques is its temporal resolution, which allows processing EEG responses in millisecond accuracy. One disadvantage of EEG to the other approaches is its spatial resolution, which is constrained, for example, by a low signal-to-noise ratio and limited spatial sampling. In addition, while EEG electrodes record the signal from a particular scalp location, it is not guaranteed that the signals originate precisely from the measured areas of the skull. EEG signals may originate from overlapping areas of the brain. This can be problematic when researchers want to investigate signals originating from the lower brain, such as the basal ganglia.

In EEG, a series of electrodes are placed on a participant's scalp. The location of these electrodes and their alignment is standardized in the 10–20 system or map of electrode placement (Jasper 1958) (see Fig. 4.2, and also the 10–10 EEG sensor placement system (Chatrian et al. 1988)), which is a reference map used to align electrodes to scalp locations corresponding to brain lobe areas. Often an electrode cap makes the placement on the head easier. EEG measures slight electrical activity, such as the signals generated by neural activity in the brain (Pizzagalli 2007). There are a variety of different measurement devices available for this type of physiological measure, ranging from more sophisticated medical grade headcap setups with high-density electrode arrays (from 32 to 256 electrodes) to simpler devices that have fewer electrodes and less spatial accuracy, but similar temporal signal accura-

[2] In the context of brain activity, the hemodynamic response consists of delivering blood to active brain tissues. The blood delivers oxygen and glucose, which are needed for brain tissue to function.

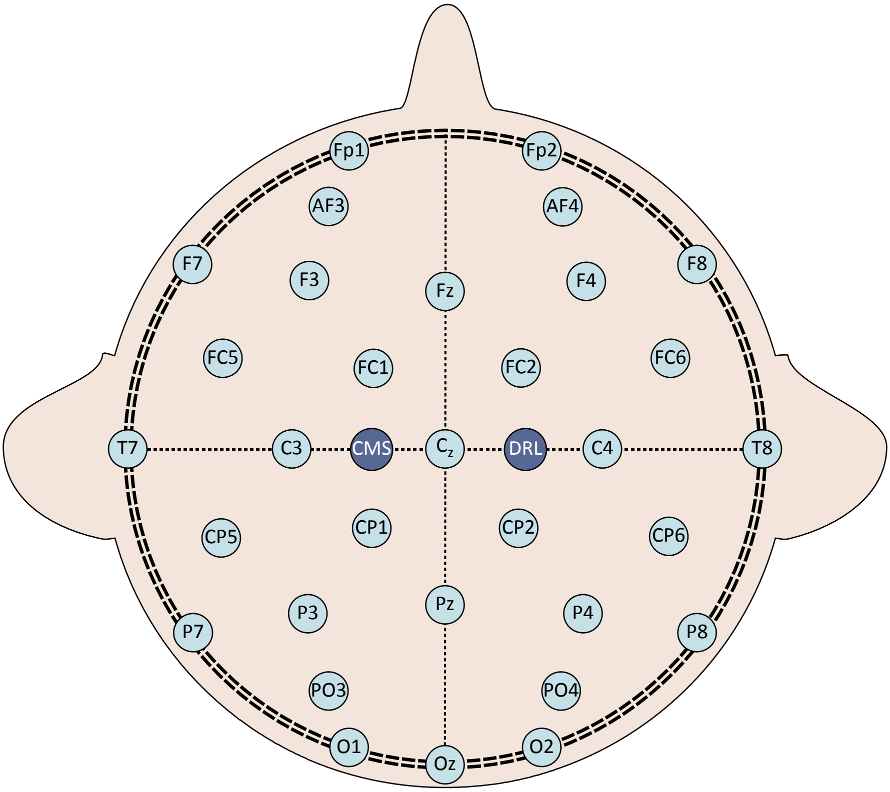


**Fig. 4.2** An example of the 10–20 electrode placement

cy. The 10-20 system indicates brain areas at scalp locations by lettering (Stern et al. 2001). For example, C is at the central, O is at the occipital, P is at the parietal, T is at the temporal, and F is at the frontal lobe site of the scalp. Additionally, the system uses numbers to indicate laterality (e.g., odd numbers of the left side, even numbers on the right side) and subscripts to indicate placement in relation to the midline.

EEG measures oscillations of inhibitory and excitatory brain activity (Pizzagalli 2007; Stern et al. 2001)[3]. This means that EEG lets us record electrical activity on the scalp that relates to brain activity. We usually distinguish brain activity by using the amplitude and frequency of the signal, which is what we can see in the graphical plots of brainwaves with which many readers are familiar. Amplitude describes the size of the signal, while frequency refers to the speed of signal cycles. Lower frequencies measure large synchronized amplitudes of neural activity, and higher

[3] The electrical potential of a neuron in the brain is influenced by neurotransmitters that either increase (e.g., excitatory signal) or decrease (e.g., inhibitory signal) the probability of producing an action potential (e.g., a short neural event during which a cell's electrical potential briefly rises and falls) in a connecting cell.

**Fig. 4.3** EEG, EMG, and GSR electrode setup shown on a study participant

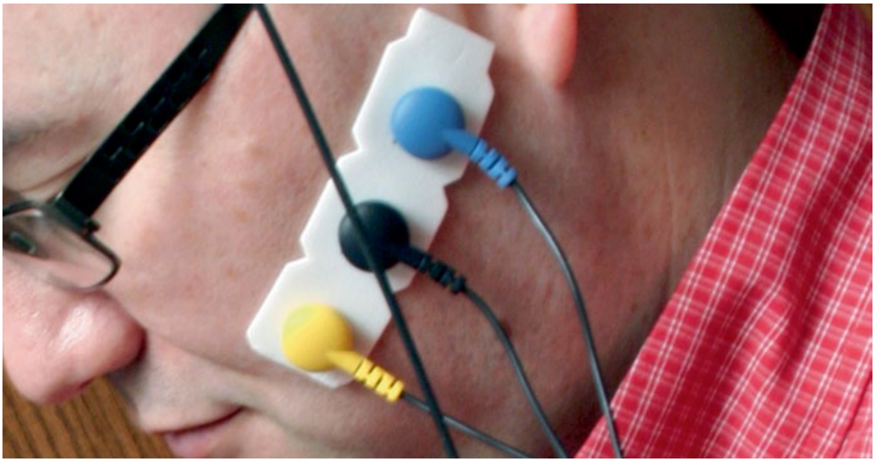

**Fig. 4.4** Example of EMG electrodes (silver-silver chloride) attached to the face of a player

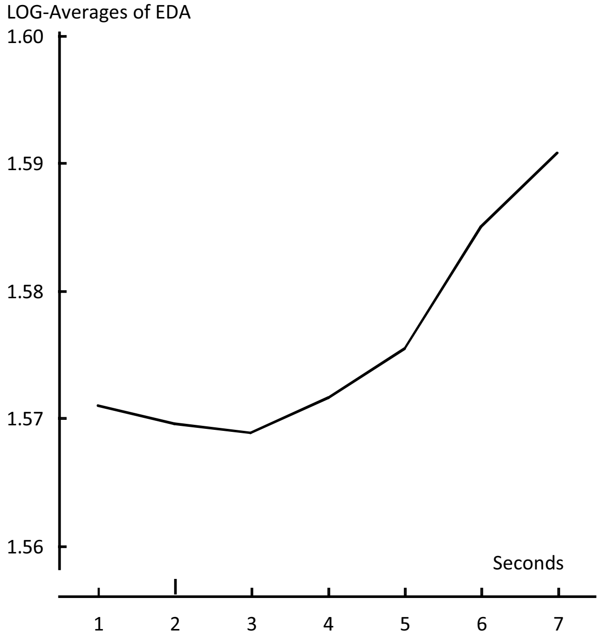


**Fig. 4.5** An example of log-normalized EDA measurement. See also the study in Nacke (2013)

frequencies show smaller, desynchronized amplitudes of neural activity. EEG measures the amplitude of this activity in dimensions of a few microvolts (µV), usually between 10 and 100 µV. This measurement consists of amplifying a differential signal between two electrodes, one at the location of measurement and others at a location of reference. Common reference points are the nose, scalp center, or mastoids. Alternatively, the average signal can be calculated as a global reference. Each electrode's electrical activity is then subtracted from the reference point.

Research-grade EEG devices compute brain waves in different frequency bands, such as alpha (e.g. 8–13 Hz), beta (e.g. 13–30 Hz), theta (e.g. 4–8 Hz), delta (1–4 Hz), and sometimes gamma (30–50 Hz)[4]. The following are some rough distinctions of frequency bands according to some of the recent literature. Alpha activity indicates a drowsy or relaxed state, and a lack of active cognitive processes; it has also been tied to information and visual processing. Beta activity has been discussed as replacing alpha rhythms during cognitive activity and is related to alertness, attention, vigilance, and excitatory problem solving activities. Theta activity has been related to decreased alertness and lower information processing. However, frontal midline theta activity in the anterior cingulate cortex scalp area is linked to mental effort, attention, and stimulus processing. This means that there are two types of theta activity: Low alertness (leading to sleep) and selective cognitive inactivity because of automated processing. Delta is most prominent during sleep, relaxation

[4] Another way of analysing EEG is through Event-Related Potentials, which I do not cover in this chapter. A more in-depth discussion of this can be found in Wehbe et al. (2013).

or fatigue, and is diminished with increasing age. Gamma activity is still largely unexplored (although recent studies relate it to the creation of meaning from separate pieces). While it has to be noted that most of these findings come from research in the medical and psychological domain, some of these simplified interpretations might make it easier to evaluate a game based on the brainwave activity revealed by EEG. For example, if increased beta activity is prevalent during gaming, it could be linked to player attention and increased arousal during a focused task. More studies are needed in this area to investigate a possible relation between this phenomenon and meaningful decision-making in games.

EEG is rather prone to artifacts, especially caused by head movement (the electrodes might move on the scalp while the player is moving). Therefore, some games are less suited for this type of evaluation (e.g., Guitar Hero, Kinect, or Wii Games that involve much movement) than others. Movement artifacts are a problem of all physiological measures, but are especially problematic with EEG, as we are interpreting very low-level electromagnetic activity. Researcher often remove artifacts by visual inspection, during which the data is inspected for irregular signals, peaks, rectangular patterns associated with lateral eye movements, or other patterns.

In addition, as with all physiological measures, EEG measures should be recorded with a preceding baseline, meaning a neutral moment (where the player does nothing but stare at a cross on a grey background, for example). This will allow any filtering procedure or algorithm applied for the analysis to filter out the signal noise that individuals naturally produce when not engaged in an activity. Since we all have different physiological activity due to natural physical differences, it is important to correct for these individual differences and focus on the changes from the normal activity in any following analysis. A final major problem with EEG as a measurement method is the difficult interpretation of the data. Similar to many physiological measures, this is called the one-to-many relationship between physiologically visible signals and the underlying psychological causes. For example, when delta activity is increased during a playing session, do we argue that the game is relaxing, or that it is boring and fatigue-inducing? It is quite important to keep one's game design goals in mind when doing this type of evaluation, and even then the process of triangulating this data with other measures is paramount for a solid interpretation.

In Fig. 4.3, an example physiological participant setup is shown. The EEG cap has all electrodes placed in locations around the scalp. Since EMG and EEG are fed through the same system, this acts as reference for the EMG electrodes as well, which allows for very small EMG sensors to be attached to the participant's face. Some additional surgical tape is used to make sure that the electrode cables do not move around too much, to mitigate movement artifacts in the physiological data.

### 4.3.6 Ten ways in Which Physiological Evaluation is Valuable to the Games Industry

With the wealth of techniques and approaches available, it is important to understand the specific benefits of physiological evaluation and games user research in the game development industry. Fairclough (2011) suggested a thought experiment, where he outlines ten suggestions for improving the use of physiological metrics in games user research. Many of these suggestions are necessary to be implemented by games user researchers for physiological metrics to work effectively in an industry setting.

1. Physiological metrics can be recorded continuously during a games user research session without interrupting play. This makes these methods less obtrusive than subjective measures that either break the experience (by interrupting and prompting with questions) or introduce memory bias (by asking questions about the game in retrospect). The only downside of physiological metrics is that the player has to wear sensors, and that some might find this intrusive (although based on personal experience, many players forget that they are wearing sensors a few minutes into the game).
2. A games user researcher interested in physiological assessment of players needs to be well-informed about what each sensor type measures. Company executives and the marketing department need to understand that this is no emotion quantifier or thought printer. Sensors measure electrical activity that can be elicited from motor, skin, or brain activity, and depending on the area of application, provide some measurement and insight as to the activity of the body area being measured. This also means that we need an experience vocabulary working from a high-level psychological concept (engagement) toward the low-level body response (sympathetic activation  higher heart rate). Inferences made from low-level body responses to high-level concepts in comparison always face challenges in withstanding closer scientific inspection.
3. For capturing player experience, a hypothesis-driven approach is suggested, where only one particular aspect of experience (ideally, one that is well-defined in the literature) is under investigation. For example, we may choose to investigate only the positive and negative emotional responses to a certain game event or game area, or the cognitive workload experienced during a game tutorial.
4. To establish a link between ideal player experience and the corresponding physiological responses, we should investigate responses to key aspects when naïve participants play the most successful games of the industry (in terms of financial and critical success). If we could find out which at physiological responses relate to the player experiences that drive the success of these games, we could work towards establishing a physiological success metric.
5. In every aspect of physiological experimentation, we need to be aware that the human body is still present in the real world while playing a video game.

Our nervous system, therefore, also responds to stimuli in this world as well as internal cognitive processes that we might have given our current environmental/living situation. These contextual influences (that may be nascent during the screening of a participant) may result in changes in emotion or motivation during the experiment. Influences such as room temperature, movement, drugs, chemicals, noise, and many more can also introduce contextual bias into our interpretation of physiological activity. In the end, it is important to keep in mind how sensitive our nervous system really is when interpreting physiological metrics.

6. Physiological metrics do not distinguish between physical activity and psychological events. Three components are involved in recording physiological metrics: external physical activity, internal emotional activity, and internal cognitive activity (Stemmler et al. 2007).
7. Given what we now know about physiological responses, we will always have a certain signal-to-noise ratio in our physiological metrics. We can counteract the amount of noise by enforcing a strict experimental protocol in a very controlled environment, or by recording all possible confounds with additional sensors (e.g., temperature, noise, light) to remove their influence during analysis.
8. Before testing players, it is important to carefully record their demographic background, including their skill level and past game preferences and experiences. Novelty and habituation can impact physiological responses considerably.
9. It is important to create the different experimental conditions carefully within a systematically manipulated environment (e.g., a game engine). Ideally, only change one variable at a time (although this is often not possible in an industry environment).
10. Other gameplay tracking metrics can be considered overt behavior markers in the game world, as they are visible instantly, whereas physiological metrics are covert measures that are not always visible directly. Both metrics should be tracked together, and a possible relationship between them should be explored using statistical analyses. Subjective responses should be recorded before and after, rather than during, physiological measurement.

We can conclude that psychophysiological measures in games should not be used alone, but always in conjunction with other measures to establish relationships between player experience facets and physiological responses. Much work remains to be done in this area before it becomes part of the everyday testing practices of games user researchers, but given recent advances by sensor manufacturers, this technology will soon become common in games user research. When using these technologies, we must be careful to remember the associated contextual influences and sensitivities, ensuring that data is obtained and analyzed in the most accurate way possible. Through precise application and exploration of this technology, we can strive to improve not only our own games, but the very state of the art in games user research.

**Acknowledgements** This research was supported by the Network of Centres of Excellence (NCE), Graphics, Animation and New Media (GRAND) and NSERC. Dr. Nacke thanks Samantha Stahlke and Rina Wehbe for proofreading the final manuscript.

## References

Ambinder M (2011) Biofeedback in gameplay: how valve measures physiology to enhance gaming experience. Paper presented at the game developers conference

Bartle R (1996) Hearts, clubs, diamonds, spades: players who suit MUDs. J MUD Res 1(1) 19 pages. http://mud.co.uk/richard/hcds.htm

Boucsein W (1992) Electrodermal activity. Plenum Press, New York

Bradley MM, Codispoti M, Cuthbert BN, Lang PJ (2001) Emotion and motivation I: defensive and appetitive reactions in picture processing. Emotion 1(3):276–298

Brathwaite B, Schreiber I (2008) Challenges for game designers. Charles River Media, Boston

Brockmyer JH, Fox CM, Curtiss KA, McBroom E, Burkhart KM, Pidruzny JN (2009) The development of the game engagement questionnaire: a measure of engagement in video game-playing. J Exp Soc Psychol 45(4):624–634. doi:10.1016/j.jesp.2009.02.016

Buss AH, Perry MP (1992) The aggression questionnaire. J Pers Soc Psychol 63(3):452–459

Cacioppo JT, Tassinary LG, Fridlund AJ (1990) The skeletomotor system. In: Cacioppo JT, Tassinary LG (eds) Principles of psychophysiology: physical, social, and inferential elements (pp 325–384). Cambridge University Press, New York, xiii, 914 pp.

Cacioppo JT, Tassinary LG, Berntson GG (2007) Psychophysiological science. In: Cacioppo JT, Tassinary LG, Berntson GG (eds) Handbook of psychophysiology. pp 3–26. Cambridge University Press. 3rd Edition. ISBN: 0521844711

Carnagey NL, Anderson CA, Bushman BJ (2007) The effect of video game violence on physiological desensitization to real-life violence. J Exp Soc Psychol 43(3):489–496

Carver CS, White TL (1994) Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. J Pers Soc Psychol 67(2):319–333

Chatrian GE, Lettich E, Nelson PL (1988) Modified nomenclature for the "10 %" electrode system. J Clin Neurophysiol: Off Publ Am Electroencephalogr Soc 5(2):183–186

Damasio AR (1994) Descartes' error. G.P. Putnam, New York

Desurvire H, Wiberg C (2008) Master of the game: assessing approachability in future game design. In: CHI '08 extended abstracts on human factors in computing systems, Florence, Italy, April 5–10, 2008. ACM, New York, pp 3177–3182. doi:10.1145/1358628.1358827.

Desurvire H, Caplan M, Toth JA (2004) Using heuristics to evaluate the playability of games. In: CHI '04 extended abstracts on Human factors in computing systems, Vienna, Austria, 2004. ACM, pp 1509–1512. doi:10.1145/985921.986102

Duchowski AT (2007) Eye tracking methodology: theory and practice, 2nd edn. Springer, New York

Eysenck SBG, Eysenck HJ, Barrett P (1985) A revised version of the psychoticism scale. Pers Individ Differ 6(1):21–29

Fairclough SH (2011) Biometrics and evaluation of gaming experience part two: a thought experiment. http://www.physiologicalcomputing.net/?p=1760. Last accessed 20 Oct 2012

Fridlund AJ, Cacioppo JT (1986) Guidelines for human electromyographic research. Psychophysiology 23(5):567–589. doi:10.1111/j.1469-8986.1986.tb00676.x

Hazlett RL (2006) Measuring emotional valence during interactive experiences: boys at video game play. Paper presented at the Proceedings of the SIGCHI conference on Human Factors in computing systems, Montréal, Québec, Canada. doi:10.1145/1124772.1124925

Jasper HH (1958) Report of the committee on methods of clinical examination in electroencephalography. Electroencephalography and Clinical Neurophysiology 10

Jegers K (2008) Investigating the Applicability of Usability and Playability Heuristics for Evaluation of Pervasive Games. In: Internet and Web Applications and Services, 2008. ICIW '08. Third International Conference on, 2008. pp 656–661. doi:10.1109/ICIW.2008.54.

Jennett C, Cox AL, Cairns P, Dhoparee S, Epps A, Tijs T, Walton A (2008) Measuring and defining the experience of immersion in games. Int J Hum Comput Stud 66:641–661. doi:10.1016/j.ijhcs.2008.04.004

John OP, Srivastava S (1999) The big five trait taxonomy: history, measurement, and theoretical perspectives. In: Pervin LA, John OP (eds) Handbook of personality: theory and research. 2nd edn. Guilford Press, New York, pp 102–138

Kivikangas JM, Ekman I, Chanel G, Järvelä S, Salminen M, Cowley B, Henttonen P, Ravaja N (2010) Review on psychophysiological methods in game research. Paper presented at the Proceedings of 1st Nordic DiGRA

Koeffel C, Hochleitner W, Leitner J, Haller M, Geven A, Tscheligi M (2010) Using heuristics to evaluate the overall user experience of video games and advanced interaction games evaluating user experience in games. In: Bernhaupt R (ed). Human–computer interaction series. Springer, London, pp 233–256. doi:10.1007/978-1-84882-963-313

Korhonen H, Koivisto EMI (2006) Playability heuristics for mobile games. In: Proceedings of the 8th conference on Human-computer interaction with mobile devices and services, Espoo, Finland, 2006. ACM, pp 9–16. doi:10.1145/1306813.1306828.

Korhonen H, Koivisto EMI (2007) Playability Heuristics for Mobile Multi-player Games. In: International conference on Digital interactive media in entertainment and arts (DIMEA), Perth, Australia, 2007. ACM, pp 28–35 doi:10.1145/1306813.1306828

Lang PJ (1995) The emotion probe. Studies of motivation and attention. Am Psychol 50:372–385

Larsen RJ, Diener E (1992) Promises and problems with the circumplex model of emotion. Rev pers soc psychol 13:25–59

Larsen JT, McGraw AP, Cacioppo JT (2001) Can people feel happy and sad at the same time? J Pers Soc Psychol 81(4):684–696. doi:10.1037/0022-3514.81.4.684

LeDoux J (1998) The emotional brain. Orion Publishing Group, London

Lewis C (1982) Using the thinking-aloud method in cognitive interface design. Vol technical report. IBM TJ Watson Research Center, Yorktown Heights

Likert R (1932) A technique for the measurement of attitudes. Arch Psychol 22(140):1–55

Livingston IJ, Mandryk RL, Stanley KG (2010) Critic-proofing: how using critic reviews and game genres can refine heuristic evaluations. Paper presented at the Proceedings of the International Academic Conference on the Future of Game Design and Technology, Vancouver, British Columbia, Canada doi:10.1145/1920778.1920786

Lykken DT, Venables PH (1971) Direct measurement of skin conductance: a proposal for standardization. Psychophysiology 8(5):656–672. doi:10.1111/j.1469-8986.1971.tb00501.x

Mandryk R (2008) Physiological measures for game evaluation. In: Isbister K, Schaffer N (eds) Game usability: advice from the experts for advancing the player experience. Morgan Kaufmann, Burlington, pp 207–235

Mandryk RL, Inkpen KM, Calvert TW (2006) using psychophysiological techniques to measure user experience with entertainment technologies. Behav Inform Technol 25(2):141–158. doi:10.1080/01449290500331156

Mirza-Babaei P, Nacke LE, Gregory J, Collins N, Fitzpatrick G (2013) How does it play better?: exploring user testing and biometric storyboards in games user research. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13). ACM, New York, NY, USA, 1499–1508. doi:10.1145/2470654.2466200. http://doi.acm.org/10.1145/2470654.2466200

Nacke L (2009) Affective ludology: scientific measurement of user experience in interactive entertainment. Ph. D. thesis, Blekinge Institute of Technology, Karlskrona

Nacke L (2010) Wiimote vs. Controller: Electroencephalographic Measurement of Affective Gameplay Interaction. Paper presented at the Proceedings of Future Play 2010, Vancouver, BC

Nacke L, Lindley CA (2008) Flow and Immersion in First-Person Shooters: Measuring the player's gameplay experience. Paper presented at the Proceedings of the 2008 Conference on Future Play: Research, Play, Share, Toronto, Canada, November 3–5 doi:10.1145/1496984.1496998

Nacke LE, Bateman C, Mandryk RL (2014) BrainHex: A Neurobiological Gamer Typology Survey. Entertain Comput 5:55–62. doi:10.1016/j.entcom.2013.06.002

Nacke LE, Kalyn M, Lough C, Mandryk RL (2011) Biofeedback game design: using direct and indirect physiological control to enhance game interaction. Paper presented at the CHI 2011, Vancouver, BC, Canada. doi:10.1145/1978942.1978958

Nacke LE (2013) An introduction to physiological player metrics for evaluating games. In: Seif El-Nasr M, Drachen A, Canossa A (eds) Game Analytics - Maximizing the Value of Player Data. Springer London, 585–619. doi:10.1007/978-1-4471-4769-5_26

Nielsen J, Molich R (1990) Heuristic evaluation of user interfaces. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people, Seattle, Washington, United States doi:10.1145/97243.97281

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant Jack L (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Curr Biol CB 21(19): 1641–1646 doi:10.1016/j.cub.2011.08.031

Pessoa L (2008) On the relationship between emotion and cognition. Nat Rev Neurosci 9(2):148–158 doi:10.1038/nrn2317

Pinelle D, Wong N, Stach T (2008a) Heuristic evaluation for games: usability principles for video game design. In: The 26th annual CHI conference on human factors in computing systems, Florence, Italy. ACM, pp 1453–1462 doi:10.1145/1357054.1357282

Pinelle D, Wong N, Stach T (2008b) Using genres to customize usability evaluations of video games. In: 2008 Conference on future play: research, play, share, Toronto, Ontario, Canada. ACM, pp 129–136 doi:10.1145/1496984.1497006

Pizzagalli DA (2007) Electroencephalography and high-density electrophysiological source localization. In: Cacioppo JT, Tassinary LG, Berntson GG (eds) Handbook of psychophysiology, 3rd edn. Cambridge University Press, New York, pp 56–84

Poels K, et al. (2010) Digital games, the Aftermath: qualitative insights into postgame experiences. Evaluating user experience in games. Springer London, 149–163. doi:10.1007/978-1-84882-963-3_9

Ravaja N, Turpeinen M, Saari T, Puttonen S, Keltikangas-Järvinen L (2008) The psychophysiology of James Bond: phasic emotional responses to violent video game events. Emotion 8(1):114–120. doi:10.1037/1528-3542.8.1.114

Russell JA (1980) A circumplex model of affect. J Pers Soc Psychol 39(6):1161–1178. doi:10.1037/h0077714

El-Nasr M, Drachen A, Canossa A (2013) Game analytics: Maximizing the value of player data. Springer London. doi:10.1007/978-1-4471-4769-5

Stemmler G, Aue T, Wacker J (2007) Anger and fear: separable effects of emotion and motivational direction on somatovisceral responses. Int J Psychophysiol 66(2):141–153. doi:10.1016/j.ijpsycho.2007.03.019

Stern RM, Ray WJ, Quigley KS (2001) Psychophysiological recording, 2nd edn. Oxford University Press, New York

van Abeelen JHF (1964) Mouse mutants studied by means of ethological methods. Genetica 34(1):79–94. doi:10.1007/bf01664181

Wehbe RR, Nacke LE (2013) An introduction to EEG analysis techniques and brain-computer interfaces for games user researchers. Proceedings of DiGRA 2013. DiGRA, Atlanta, GA, United States, pp 1–16

White GR, Mirza-Babaei P, McAllister G, Good J (2011) Weak inter-rater reliability in heuristic evaluation of video games. Paper presented at the Proceedings of CHI EA 2011, Vancouver, BC, Canada doi:10.1145/1979742.1979788