A Proposal for the Selection of Eye-Tracking Metrics for the Implementation of Adaptive Gameplay in Virtual Reality Based Games

Jose L. Soler-Dominguez[1], Jorge D. Camba[2], Manuel Contero[1(✉)], and Mariano Alcañiz[1]

[1] Instituto de Investigación e Innovación en Bioingeniería (I3B),
Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain
jlsoler@lableni.com, mcontero@upv.es, malcaniz@i3b.upv.es
[2] Gerald D. Hines College of Architecture and Design, University of Houston,
4200 Elgin St., Houston, TX 77204-4000, USA
jdorribo@uh.edu

Abstract. Eye tracking has been shown to be an effective method for evaluating usability and User eXperience (UX) in various environments. With the advent of affordable high performance Head Mounted Displays (HMDs), Virtual Reality (VR) has rapidly expanded and opened new channels for fun and "serious" game experiences. However, there is a lack of effective quality indicators that measure satisfaction, frustration, boredom, or attention of users immersed in VR environments. Traditional approaches such as questionnaires, surveys, and interviews can easily interfere with the user's sense of presence if the assessment is performed while the experience is in progress. If performed after completing the experience, users are forced to recall memories that may be incomplete or inaccurate. In this context, the integration of eye tracking technology into HMDs provides a reliable and non-intrusive mechanism to assess UX and feed real-time data which can be used to design and develop adaptive VR experiences. In this paper, we discuss the assessment capabilities of eye tracking technologies in immersive VR environments and describe strategies to apply this information to the creation of VR games that can adapt to the player's performance, mood and behavior.

Keywords: Eye tracking · Virtual reality · Games · Serious games · Adaptive game design

1 Introduction

Game developers make great efforts to avoid players' frustration while keeping engagement with the gameplay. This characteristic of modern games contrasts with classic 80s and 90s games, where fewer levels and smaller more difficult scenarios and worlds were common. In particular, AAA games (big productions developed by large studios and aiming to be global bestsellers) and Free-to-Play (F2P) games (games where players can access a significant portion of the

content at no cost but offer in-game purchases to access additional content or special items) strive to keep players engaged during longer periods of time by implementing adaptive gameplay techniques, where the gaming experience can be catered to each individual player.

The adaptive capabilities of games are often limited to a small subset of features within the gameplay. For example, they could include leveraging Non Playable Characters' (NPCs) behavior to the player's performance, or adjusting the speed of enemies or hazards in a particular level based on the player's abilities (faster if the player is performing well, or slower if he or she is consistently failing to advance). As described by [4], NPCs in current games are equipped with the most basic operational behavior that restricts their capabilities and movements. For example, an action initiated by the user can be misinterpreted, or even missed entirely by NPCs if it falls outside the scope of their simple programmed logic.

The use of adaptive gameplay could enable interaction with the user at a higher level by predicting the intentions of the user based on previously recognised sequences of actions, rather than simply responding to low-level actions. Furthermore, adaptability is expected to grow over low-level actions or even player performance, because user emotion or how the player is feeling at a particular moment can be automatically detected. Why not change the style or difficulty of the gameplay when players are frustrated? New waves of enemies could be automatically released when the system detects the player is getting bored; soothing music can be played and a slower pace gameplay could be presented if unreasonably high stress levels are suddenly identified. In virtual reality environments, this type of experience customization is even more relevant, as presence (the feeling of being inside the virtual world) could be negatively affected by "one size fits all" game design.

Being able to detect and measure the player's feelings, emotions, and performance in real time and in a non-intrusive manner can help us anticipate the player's actions and understand how the loss of immersion can be minimized or even prevented. This information becomes fundamental in serious games, as players need to be kept in the flow channel (not anxious or bored, but focused on a particular activity) in order to have enough cognitive capacity to learn what they are expected to [5]. For games, the factors involved in making a player stay in the flow channel where described by [5]. These factors include: challenges vs. skills, anxiety vs. boredom, and difficulty balance, as shown in Fig. 1.

Many methods and tools exist to detect arousal, frustration, and other psychological states in a person. For the last 30 years, eye tracking technology has demonstrated great performance and robustness in detecting these states. Commercially, the integration of eye tracking technology in HMDs will be launched in 2017 in products such as the HTC-Vive/Tobii bundle or the FOVE headset. These products have a great potential to facilitate the use of eye-tracking metrics in the implementation of adaptive gameplay.
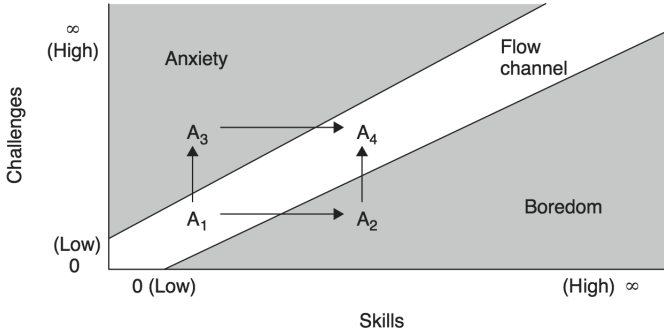
Fig. 1. Flow channel, from Jesse Schell's Art of Game Design: A Book of Lenses. "A" represents the player [5].

2 Eye Tracking Basics and Evolution

Eye tracking is the process of measuring eye positions (where a person is looking) or eye movement (the motion of an eye relative to the head). The process has become increasingly popular in Human-Computer Interaction, usability, and UX studies as a method to record the activity of the eyes and study how a person responds to visual stimuli.

The first eye tracking devices were invasive, usually based on electro oculographic systems [2] where electrodes attached to the skin around the eye detect variations in electric potential caused by eye movements; or used large contact lenses that often covered the cornea and sclera [3]. Modern eye tracking systems use video images of the eye to determine where the individual is looking (i.e., the "point of regard" [3]). There are different methods to calculate the point of regard. In commercial systems, the most commonly used method is the video-based combined pupil/corneal reflection technique. To provide this metric, the head must be fixed so that the eye's position relative to the head and point of regard coincide. A more comfortable option to disambiguate head movement from eye rotation involves combining different ocular features such as the corneal reflection of a light source (usually infra-red) and the pupil center [3], as shown in Fig. 2

The corneal reflection of the infra-red light is measured relative to the location of the pupil's center. The corneal reflections are called Purkinje reflections, Purkinje reflexes, or Purkinje-Sanson images. Purkinje-Sanson images are four defined reflections whose positions and intensities inform us about where a person is looking (see Fig. 3) [7].

The exact algorithms, sensors, and illumination technologies used by commercial eye trackers are often proprietary as these elements determine to a great extent the robustness, performance, reliability, and accuracy of the system.
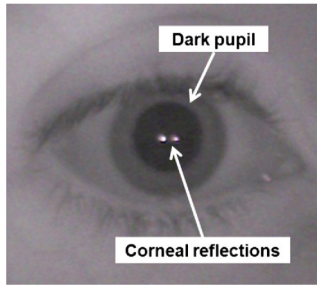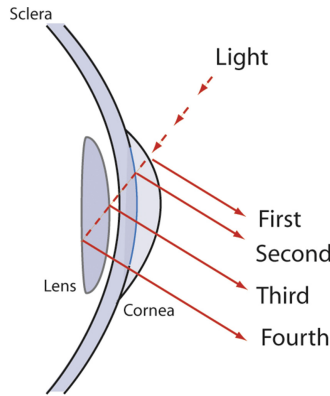
Fig. 2. Corneal reflection [6]

Fig. 3. Purkinje reflections [8]

3   Eye Tracking and HCI

Where and how an individual looks at a particular visual stimulus can provide valuable information (both conscious and unconscious) about the person's feelings and intentions. More specifically, this information can help us understand attention patterns and other high level cognitive processes [9].

Over the years, researchers have established a number of eye tracking metrics and interpretations for these metrics. The most significant ones are discussed in the following subsections based on the compilation prepared by [3]. We highlight the most relevant ones for assessing UX in games and serious games.

3.1   Eye-Movement Metrics

The most commonly used metrics in eye-tracking research include fixations (period where the eyes focus and lock towards an object) and saccades (quick eye movements between fixations).

Fixations: The meaning and interpretation of fixations can vary slightly depending on the context. Higher *fixation frequency* on an Area of Interest (AoI) in an encoding task could represent a particular interest in the target or that the target is complex and difficult to encode. In other situations such as searching or looking for information, these interpretations may be different [10]. Relevant measures derived from fixations include:

– *Fixation duration*, linked to the processing time applied to the object being fixated [11].
– *Gaze*, normally defined as the sum of all fixation durations within a prescribed area. It is typically used to compare attention between different elements. It can also be used as an anticipation metric if longer gazes occur in an AoI expecting a possible event [12].
– *Time to first fixation on-target*, defined as the amount of time that it takes a person to look at a specific AoI from stimulus onset. The sooner a person looks at a specific target, the better attention-getting properties the target is considered to have [13].

Saccades: Represent fast eye movements between fixations [14]. During saccades, any type of processing activity can occur, so they do not inform us about the complexity or prominence of an item inside a virtual environment. However, backtracking eye-movements or "regressive saccades" can give us important information about the difficulty of an encoding task. For instance, even smaller regressions in a reading task (two or three letters back) can represent confusion in the higher-level processing of text [2]. A relevant subset of saccade-derived measures include:

– *Number of saccades.* The higher the number of saccades, the more active the searching activity [15].
– *Saccade amplitude* represents a clear indicator of attention: larger saccades reveal more meaningful visual elements (requesting user focus from distance).

Scanpaths: Completion of a full cycle of saccade-fixation-saccade. The most significant metrics derived from this element include:

– *Scanpath duration:* Longer scanpaths imply a less efficient scanning of the environment [15].
– *Saccade/fixation ratio* compares the time spent searching (saccades) to the time spent processing (fixating) [3,15].

3.2 Non Eye-Movement Metrics

According to the cognitive theory of multimedia learning established by Mayer [16], multimedia naturally supports the way the human brain learns. The multimedia principle states that words and graphics are more conducive to learning than words or graphics alone [16]. The theory was summarized by [17] as

having the following components: (a) a dual structure of visual and auditory channels, (b) limited processing capacity in memory, (c) three memory storage (sensory, working, long-term), (d) five cognitive processes of selecting, organizing, and integrating (selecting words, selecting images, organizing work, organizing images, and integrating new knowledge with prior knowledge), as well as theory-grounded and evidence-based multimedia instructional methods. The implications of item (b) are particularly relevant. If humans use too much of our limited capacity trying to interact with the environment, learning the rules, understanding the tasks we are supposed to perform, and reading menus and other interfaces in VR environments, we will only have a small amount of processing capacity left to store new knowledge.

Two non eye-movement metrics are relevant: *Blink rate* and *pupil size*. Both metrics can be used as indicators of cognitive load. A low blink rate is assumed to indicate a higher cognitive workload whereas a high blink rate may indicate fatigue [18]. Additionally, a higher cognitive load is also associated to larger pupils [19].

However, non eye-movement metrics are not completely reliable, as pupil size and blink rate can be affected by external factors [20] such as the lighting conditions of the VR environment or the brightness settings in the HMD. Consequently, these metrics are not extensively used in eye-tracking research. Nevertheless, their reliability and accuracy could be significantly improved with proper calibration, where the average pupil sizes and blink rates under different lighting conditions would serve as baselines for interpretation.

4   Evaluating UX

In the gaming domain, collecting and analyzing high quality, precise, and objective data about the player is a crucial step towards understanding the overall experience. In this paper, we review some of the methods used to gather information about a player's emotional and cognitive processes, intentionally omitting those methods that are subjective such as direct observation and self-reporting. In this regard, physiological processes are mostly involuntary, so measurements are not biased by the person's answering patterns, social desirability, interpretations of wording, limits in participant memory, or observer bias [1].

Physical reactions are part of the processes that contribute to the player's game experience. However, the relationships between physiological processes and elements are not one-to-one, which makes them difficult to isolate. By analyzing the objective methods proposed by Nacke et al. [14] to assess Gameplay Experience, two groups are identified:

– Psychophysiological player testing, with technologies such as:
  • Electromyography (EMG): Technology for measuring the electrical activity of muscles. It is an effective metric for basic emotions (e.g., facial muscles are good indicators of basic emotional states).

- Electrodermal activity (EDA): Common psycho-physiological method that measures variation of the electrical properties of the skin in response to sweat, which is directly related to physical arousal.
- Electroencephalography (EEG): Technique to measure brain activity (typically using scalp electrodes).
- Functional magnetic resonance imaging (fMRI) and positron emission tomography (PET): other techniques for measuring brain activity. Both methods have strong equipment limitations, which makes them difficult to use. Functional near-infrared spectroscopy (fNIR) has recently gained significant attention among UX researchers.
- Eyetracking: Technique based on measuring saccades (fast movements) and fixations (dwell times) of human gaze [15], as explained in the previous section. Psycho-physiological measures are not incompatible with eye tracking, but complementary, as they provide additional information about the interaction with the VR environment.

5   Adaptive Gameplay

The metrics described in previous sections provide a toolbox for designers to choose from, so UX gaming experiences can be customized based on the player's behavior. By grouping the areas with potential for player adaptation, three dimensions can be defined: perceptual adaptation (adaptation of how players see the VR world), cognitive adaptation (adaptation of how the user interface or the game mechanics are conditioned by the player's performance), and affective adaptation (adaptation of the game narrative to the player's mood).

5.1   Perceptual Adaptation

The design of a gaze-contingent system must distinguish the characteristics of foveal and peripheral vision. The human eye can see approximately 135° vertically and 160° horizontally, but with clear and fine resolution only around a 5° circle located in the center. This small slice of the visual field projects to the retinal region called the fovea, which is full of color cone photoreceptors. The angular distance away from the central gaze direction is called eccentricity. Acuity falls off rapidly as eccentricity increases due to a reduced receptor and ganglion density in the retina, a reduced optical nerve "bandwidth", and a reduced "processing" devoted to the periphery in the visual cortex [23]. According to Strasburger [24], current computer graphics (CG) researchers use the term foveation as a shorthand for the decrease in acuity with eccentricity in the human visual system.

Today, CG practice ignores user point-of-regard and always renders a full high-resolution image on the entire display (screen or HMD), which is a waste of computational resources [23]. If we were able to only render a high resolution image of the 5° foveal region (approx. 0.8% of the solid angle of a 60° display [23]) while keeping the rest of the image at lower resolution, we could
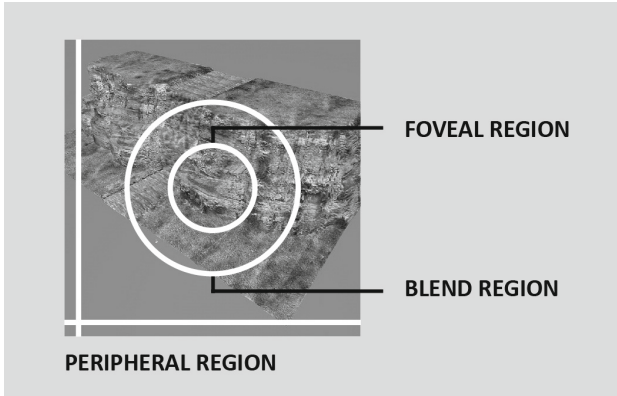
Fig. 4. Foveated rendering scheme

deploy immersive VR environments with incredible detail while saving significant graphical and computational resources. This technique is called foveated rendering (see Fig. 4) and is slowly becoming a reality in VR thanks to the advanced eye-tracking capabilities of HMDs [25]. In theory, foveated rendering could bring significant reductions in rendering cost with no discernible differences in visual quality, which is specially important in scenarios with limited computing graphic power, such as current smartphone-based VR systems. An important element of the foveated rendering technique is a high-speed gaze tracker that can prevent lags while rendering the environment and rapidly respond to the user's head movements.

5.2 Cognitive Adaptation

In the area of psycho-physiological assessment of cognitive load, the pioneering work of Goldberg and Kotval in User eXperience (UX) research is particularly relevant [15]. In their studies, the authors connected cognitive load to processing (the brain activity that deals with information, decisions, and perception). By analyzing how processing tasks affect our sight, they determined that quantity and type of fixations provide reliable metrics to measure cognitive load.

– *Number of fixations:* The higher the cognitive load, the smaller the number of fixations. Our brain cannot process searches when it is busy processing something else. In a gaming environment, if a player is permanently searching, chances are he or she is "lost" (e.g., menus may not be clear, the player does not understand the goals of a particular task, etc.). Therefore, if a significant increase in the number of saccades is observed, experience designers may have to implement strategies to clarify certain aspects of the experience such as automatically launching help messages or reminders, or providing visual cues regarding next steps for a particular task.

– *Fixation duration:* Longer fixations imply the cognitive load is higher and the user is spending more time "processing." Average fixation duration is calculated by adding the number of gaze point samples in all the fixations and dividing by the total number of fixations. This metric provides information about a player's attention and facilitates adapting the timing of certain events to the player's behavior. For example, simple actions such as repeating instructions or reinforcing visual stimuli can catch the player's attention when he or she is focused on a different game element or processing other information.

– *Fixation/saccade ratio:* This content-independent ratio compares the time spent processing (fixations) to the time spent searching (saccades). Higher ratios indicate higher cognitive loads in the VR environment. In situations where higher ratios are detected, designers should provide mechanisms to temporarily reduce the cognitive load. For example, multimedia elements in games may create unnecessary cognitive loads [27]. A simple solution could be "muting" or stopping unnecessary multimedia inputs such as audio and animations. An additional implication of high ratios is that no new information or goal should be presented to the player until the ratio has decreased, as the cognitive capacity of the player can quickly saturate.

5.3   Affective Adaptation

Humans are emotional and social creatures. We cannot completely dissociate high-level cognitive skills (e.g., reasoning, decision making, speaking, reading, or mathematics) from feelings and emotional responses. These strong links have important implications in many areas such as education and neuroscience. For example, if we could fully understand the links between emotions and behavior, it would be possible to tailor our environments to individual users and develop more effective teaching strategies [21].

Although the scope of our proposal is limited to detecting emotion arousal, identifying the specific emotion that a particular experience is generating in a user is also worth studying. In the context of our paper, we want the environment to present stimuli that can promote emotion arousal to increase engagement and motivation. Fear, for example, is an unpleasant emotion that can easily disengage users (although it could arguably be useful in specific contexts).

Several studies on the relationship between emotion and pupil change have empirically shown that pupil diameter increases when people are exposed to emotionally engaging (pleasant or unpleasant) visual stimuli, regardless of hedonic valence [22]. An important adaptation designers can provide with this metric is the introduction of coherent gameplay events that serve as emotional stimuli if a predefined emotional threshold value is exceeded. However, pupil size is not a robust metric. Thus, researchers and designers should first consider establishing baselines for pupil sizes under various conditions before implementing this metric.

The previous type of adaptation could also be linked to game narratives, as described by Cavazza [26]. Based on the author's work on the emotional input

on interactive storytelling, game designers could include branching narratives inside the main game plot that could be expanded depending on how a player emotionally reacts to the different events happening within the game.

6   Conclusions and Further Work

New technological advances in the fields of VR and eye-tracking have facilitated the development of high resolution head mounted displays with embedded eye-tracking capabilities. This new generation of VR devices allows researchers and designers to collect vasts amounts of data related to the user's mood, performance, and behavior, which can then be analyzed post-experience to inform the redesign of levels, environments, or specific gameplay aspects.

As the speed and accuracy of eye-trackers continues to increase, the real time interpretation of the metrics described in this paper and the consequent adaptation of gameplay to the particular emotional state of a player may soon become commonplace. However, it will be up to the designers to create VR environments that, even when adaptive, will remain consistent, coherent, and fully immersive (without letting the feeling of presence decrease due to artificial events triggered by the eye-tracker).

More empirical evidence is still needed regarding the robustness of the metrics and the best approaches to map these metrics to the player's emotional states. In this regard, research experiments on affective gaming rarely consider the implications of outside factors. For adaptive games, however, being able to isolate the metrics from external factors is a must in order to maintain a fully immersive experience.

Auto-calibration is also a desirable feature in adaptive eye tracking driven VR systems, as manual calibration in current eye-trackers accounts for a significant amount of preparation time in every eye-tracking experiment. Many researchers are already focusing their efforts on this topic. Finally, the relationships between game mechanics and the adaptation strategies presented in this study will also need to be defined.

References

1. Kivikangas, J.M., Chanel, G., Cowley, B., et al.: A review of the use of psychophysiological methods in game research. J. Gaming Virtual Worlds **3**(3), 181–199 (2011)
2. Rayner, K., Pollatsek, A.: The Psychology of Reading. Prentice Hall, Englewood Cliffs (1989)
3. Duchowsky, A.T.: Eye Tracking Methodology: Theory and Practice. Springer, London (2003)
4. Kazmi, S., Palmer, I.J.: Action recognition for support of adaptive gameplay: a case study of a first person shooter. Intl. J. Comput. Games Technol. **2010**, 11 (2010). doi:10.1155/2010/536480. Article 1
5. Schell, J.: The Art of Game Design: A Book of Lenses. Morgan Kaufmann Publishers Inc., San Francisco (2008)

6. Fassi, A., Riboldi, M., Forlani, C.F., Baroni, G.: Optical eye tracking system for noninvasive and automatic monitoring of eye position and movements in radiotherapy treatments of ocular tumors. Appl. Opt. **51**, 2441–2450 (2012)

7. Crane, H.D., Steele, C.M.: Generation-V dual-Purkinje Image eyetracker. Appl. Opt. **24**, 527–537 (1985)

8. Ji, Q., Hansen, D.W.: In the eye of the beholder: a survey of models for eyes and gaze. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 478–500 (2010)

9. Just, M.A., Carpenter, P.A.: A theory of reading: from eye fixations to comprehension. Psychol. Rev. **87**(4), 329 (1980)

10. Jacob, R.J.K., Karn, K.S.: Eye tracking in human-computer interaction and usability research: ready to deliver the promises. In: The Mind's Eye: Cognitive The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research, pp. 573–603 (2003)

11. Just, M.A., Carpenter, P.A.: Eye fixations and cognitive processes. Cogn. Psychol. **8**(4), 441–480 (1976)

12. Nodine, C.F., Mello-Thoms, C., Kundel, H.L., Weinstein, S.P.: Time course of perception and decision making during mammographic interpretation. Am. J. Roentgenol. **179**, 917–923 (2002)

13. Byrne, M.D., Anderson, J.R., Douglass, S., Matessa, M.: Eye tracking the visual search of click-down menus. In: Proceedings of Human Factors in Computing Systems, CHI 1999, pp. 402–409. Addison Wesley, Reading (1999)

14. Nacke, L.E., Drachen, A., Goebel, S.: Methods for evaluating gameplay experience in a serious gaming context. Intl. J. Comput. Sci. Sport **9**, 2 (2010)

15. Goldberg, J.H., Kotval, X.P.: Computer interface evaluation using eye movements: methods and constructs. Int. J. Ind. Ergon. **24**(6), 631–645 (1999)

16. Mayer, R.E.: Cognitive theory of multimedia learning. In: Mayer, R.E. (ed.) The Cambridge Handbook of Multimedia Learning. Cambridge University Press, New York (2005)

17. Sorden, S.: The cognitive theory of multimedia learning. In: Handbook of Educational Theories. Information Age Publishing, Charlotte (2012)

18. Bruneau, D., Sasse, M.A., McCarthy, J.D.: The eyes never lie: the use of eyetracking data in HCI research. In: Presented at Proceedings of the Workshop on Physilogical Computing, CHI 2002, Minneapolis, 21 April 2002

19. Pomplun, M., Sunkara, S.: Pupil dilation as an indicator of cognitive workload in human-computer interaction. In: Proceedings of the International Conference on HumanComputer Interaction (2003)

20. Goldberg, J.H., Wichansky, A.M.: Eye tracking in usability evaluation: a practitioner's guide. In: Hyönö, J., Radach, R., Deubel, H. (eds.) The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research, Amsterdam, The Netherlands (2003)

21. Immordino-Yang, M.H., Faeth, M.: The role of emotion and skilled intuition in learning. In: Mind, Brain, and Education: Neuroscience Implications for the Classroom, pp. 69–83 (2010)

22. Bradley, M.M., Miccoli, L., Escrig, M.A., Lang, P.J.: The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology **45**, 602–607 (2008)

23. Guenter, B., Finch, M., Drucker, S., Tan, D., Snyder, J.: Foveated 3D graphics. ACM Trans. Graph. **31**(6), 10 (2012). doi:10.1145/2366145.2366183. Article 164

24. Strasburger, H., Rentschler, I., Jüttner, M.: Peripheral vision and pattern recognition: a review. J. Vis. **11**(5), 13 (2011). doi:10.1167/11.5.13

25. Patney, A., Kim, J., Salvi, M., Kaplanyan, A., Wyman, C., Benty, N., Lefohn, A., Luebke, D.: Perceptually-based foveated virtual reality. In: ACM SIGGRAPH 2016 Emerging Technologies (SIGGRAPH 2016), Article 17, p. 2. ACM, New York. doi:10.1145/2929464.2929472

26. Cavazza, M., Pizzi, D., Charles, F., Vogt, T., Andre, E.: Emotional input for character-based interactive storytelling. In: Sierra, C., Castelfranchi, C., Decker, K.S., Sichman, J.S. (eds.) Proceedings of 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009), Budapest, Hungary, May 2009, pp. 313–320 (2009)

27. Kiili, K.: Digital game-based learning: towards an experiential gaming model. Internet High. Educ. **8**, 13–24 (2005)