Impact of Visual and Sound Orchestration on Physiological Arousal and Tension in a Horror Game

Sarra Graja, Phil Lopes, *Member, IEEE*, and Guillaume Chanel

***Abstract*—Horror games represent a very specific genre specifically designed to elicit fear. These games provide a tremendous emotional experience balanced between stress and satisfaction. Yet, over time, the player acquires further insight of the mechanisms of the game, dissipating the creepy climate that reduces the emotional impact intended. This article hypothesises that exploring existing facets within a game such as visuals and sounds might establish a good approach to renew the gaming experience. To understand the players' emotional reactions toward context alteration, an adaptation of a published game (*P.T.* by Konami 2014) was used. This context refers mainly to light effects, sounds, and in-game events. To learn which game effects induce the strongest physiological reactions, an experiment was conducted, and correlation between the physiological data, collected through the measure of the galvanic skin response, and that of the perceived emotion provided by participants, was investigated. Results show that the order in which effects are arranged can produce extensive emotional responses. They also suggest that psychological impact can be increased not only by the visual horror itself, but also through the process that slowly builds up to it, in particular the usage of sounds.**

***Index Terms*—Affective computing, emotion, galvanic skin response (GSR), game content orchestration, gameplay experience, human–computer interaction, player response.**

I. Introduction

VIDEO games are interactive spaces genuinely capable of eliciting complex emotional responses. This makes video games an excellent way for acquiring new insights about human emotions and the way of conceiving better human–machine interaction systems. Users resort to video games in search of new emotional experiences. Some play knowing they might face negative emotions such as fear and frustration. Video games do indeed possess a great potential in influencing players by manipulating several aspects of the virtual world. The process that gradually unbinds them from the real world and immerses them into a virtual one leads them to develop complex affective, behavioral, and cognitive reactions. In fact, emotion forms the core element that commands the process of building video games. Game designers often develop gameplaying experiences with the intent of transmitting specific interactive and/or emotional experiences to their players. Stress, frustration, anticipation, and commitment all offer a crucial perspective for the development process. As such, it makes sense to explore how virtual environmental aesthetics can be manipulated and transformed in such a way that it may convey different player emotional experiences.

Manuscript received March 10, 2020; revised May 12, 2020; accepted June 19, 2020. Date of publication June 30, 2020; date of current version September 15, 2021. This work was supported in part by Innosuisse. *(Corresponding author: Sarra Graja.)*

Sarra Graja and Guillaume Chanel are with the Social Intelligence and Multi-Sensing Group and the Computer Vision and Multimedia Laboratory, University of Geneva, CH-1227 Carouge, Switzerland (e-mail: sarra.graja@unige.ch; guillaume.chanel@unige.ch).

Phil Lopes is with the Immersive Interaction Group, École Polytécnique de Lausanne, CH-1015 Lausanne, Switzerland (e-mail: phil.lopes@epfl.ch).

Digital Object Identifier 10.1109/TG.2020.3006053

A. Motivation

In view of the community's growth in the video game industry, diversifying the content becomes crucial in order to cover players' expectations and preferences. Tailoring this content to an individual's emotional profile becomes, therefore, an increasingly difficult task. Yannakakis and Togelius [1] argued, during their research on optimization tools for gaming experience, that the holistic approach for affective synthesis in games requires the integration of the game's content into the computational model of affect. Thus, they introduced a framework for procedural content generation (PCG), driven by computational models of user experience, called experience-driven procedural content generation (EDPCG). Although experiences with the EDPCG method have been successful, games being highly immersive and interactive can lead to complex affective responses that cannot be evaluated in view of the number of possible combinations between the generated content and the affective models of playing behavior.

This article focuses on two main facets of game development, specifically visuals and sounds. The aim is to identify which effects are good stress elicitors by observing the players' physiological reactivity and then comparing it to their own perception of stress/anxiety/tension. Thus, the question is formulated as such: Which game effects can induce the strongest physiological reactions? Are these physiological reactions correlated with the players' perceived emotions?

The main purpose behind the understanding of players' emotional reactions toward context alteration is to help game designers make accurate decisions in how to orchestrate this context to have games more engaging and impactful and achieve manifold gaming experiences either by looking up the information themselves or by conceiving better algorithms.

To conduct our study, we chose to explore the concept adopted in the popular demo *P.T* (Konami 2014) through *PTVR*, a variant of the game, developed by M. Svendsen and T. Clavero (2017). Within this game, the level architecture is static, and only the context essentially relating to lights, sounds, and actions (in-game events) is altered. Furthermore, despite its name, it is no longer a virtual reality (VR) game and is played with mouse and keyboard similarly to other types of first person shooters.

II. Related Work

A. Emotion

The topic of human emotions has always been a contentious one within psychology, with several theoretical models emerging to explain the dichotomies between the different emotions perceived [2]–[4]. Furthermore, it is suggested that when a certain emotion is induced, common patterns may emerge from human bodily functions, e.g., such as facial expressions or changes in heart rate (HR), thus theoretically providing a means to "connect" emotions to their relevant human bodily reactions [5].

To classify emotion, Ekman and Friesen [2] and Tomkins [3] defined six discrete categories, namely happiness, surprise, anger, sadness, disgust, and fear, considered biologically fixed and universal for every individual. Russell's opposite thinking [4] regards emotions as combinations along the dimensions of arousal and valence, where the valence dimension contrasts the states of pleasantness (happy) with states of unpleasantness (sad), and that of the arousal contrasts the states of low activation (calm) with states of high activation (surprise).

1) Emotion and Physiology: To decode emotion, psychology relies on physiological changes induced by the autonomic nervous system (ANS). Cacioppo *et al.* [6] and Mauss and Robinson [7] proposed that "different measures of emotion appear sensitive to different dimensional aspects of state rather than distinct emotional states." Nevertheless, measuring the somatic changes induced by the ANS does not yet make it possible to detect precisely and reliably the valence level corresponding to a specific arousal value [8]. To obtain a more precise description of the different emotional state, the PAD dimensional model [9] was introduced, which combines pleasure, arousal, and dominance, though it is still recommended and indeed essential to include additional data such as surveys [10] or video analysis [11].

Within this project, we are interested in measuring tension and stress. Stress is interesting regarding its physiological and behavioral responses induced due to intensity and duration [12]. One reliable method for measuring physiological and mental stress [13] and collecting data in an objective and nonintrusive manner is the use of the measurement sensor of the galvanic skin response (GSR). GSR is, in fact, an independent index of sympathetic activity, a characteristic of negative emotions [14] that controls the sweat gland activity [15]. The secreted sweat is captured by the GSR through small electrical pulses and that often correlates with arousal [16]. This tool shows a very good accuracy in comparison with other tools such as HR, respiratory rate, or blood pressure (BP) [17]. HR, for example, is antagonistically controlled by the sympathetic activity and the parasympathetic [18] [19]. This makes GSR a more reliable measure of tension, given that it is a direct response to the acute stress response. However, as previously mentioned, the GSR sensor does not distinguish between positive and negative arousal states. For example, fear, surprise, and ecstasy are among the emotions that can produce similar responses. For data to be relevant, the environment is framed within the horror context. This study relies on effects proven to be effective elicitors of stress and anxiety (see Section II-C2). This also ensures that responses to unexpectedness correspond to startle reflexes and not to discovery or wonder/admiration often found with surprise. In fact, startle takes place under the threshold of surprise and shock and occurs on a purely passive level of experience; without any reconstitution of sense as it is the case for surprise (reconstitution of sense—"believing what I can't believe") and shock (resistance to the reconstitution of sense—"I can't believe what I can't believe") [20]. Finally, surveys and postexperimental questioning sessions are used and cross-examined alongside the empirical data.

2) Emotion in Digital Games: Game designers often incorporate emotional intention within their designs, as the emotional connections between the player and the game can lead to an enhanced experience [21]. Such concepts have often been touched upon by several researchers attempting to investigate how emotion can lead to better experiences [22]. Given that the nature of games consists of players overcoming specific challenges, the theory of flow [23] is often regarded as one of the most applicable for the field of games [22], [24]. More so this is true within the context of the horror genre, where the objective of the said game is to provide strong emotional experiences. In fact, it can be considered that one of the core challenges of such games is actually overcoming the emotion itself so as to reach the end. This work attempts to explore the concept of horror within the digital game space by looking at how the orchestration of visuals, audio, and in-game interactions can be used to induce such negative emotions.

B. Game Orchestration

Liapis *et al.* [25] distinguished six different artistic facets in games' context: visuals, audio, level architecture (level design), narrative, ludus (game rules), and gameplay (players interaction with a game). Combined together, they ensure an emergent interactive experience. Previous studies explored the concept of orchestration of some of these facets such as audio and gameplay in *AudioInSpace* [26], where the space shooter's shooting mechanics evolve based on the background music, and game rules (ludus) along with level architecture in studies with the Angelina system [27] for automatic game generation. Lopes *et al.* [28] explored the orchestration concept between audio and level architecture with the intention of creating tense and frightful gaming experiences within horror games. They developed *Sonancia*, a game which put emphasis on methods that have the capacity of successfully sonifying unseen, procedurally generated levels. Their focus was mostly on audio as it plays, in such games, an important role in establishing an uneasy mood around the player without him facing a direct visual representation. In the context of this work, we seek to know which audio and visuals are the most effective in inducing stress and anxiety. Contrary to previous cited works, we are more interested in exploring the emotional impact that results from this orchestration.

C. Importance of Emotion Within Horror

Fear is "an experience," describes A. Chamberlain (1899) [29]. In his research of the unattractiveness of the disgust in films, Mccauley [30] concluded that viewers can enjoy experiencing negative emotions such as fear or disgust. According to Carroll *et al.* [31], "horror attracts because anomalies command attention and elicit curiosity." Horror movies in their violation of norms hold a kind of fascination for people as they rarely witness these violations in everyday experience. These films satisfy spectators' curiosity, which make them immediately enjoyed. In contrast, based on Zuckerman's Sensation Seeking Scale, [32] it is assumed possible that "horror films provide the kind of intense stimulation and arousal that will appeal particularly to high sensation seekers" [33]. Another hypothesis by Rickey [33] suggest that the successful resolution of a horror film is what makes it enjoyable. King [34] himself called it "the magical moment of the reinstatement of safety at the end." The elimination of an unpleasant stimulus proves to be rewarding [35]. However, a distinction between games and movies is to be mentioned. According to Vorderer and Knobloch [36], based on Zillmann's theory of disposition, the experiential state of the viewer is likely to transform into actual fear or hope if the latter were to intervene and influence the plot. In films, viewers act as mere observers hoping for the hero to survive, whereas in horror games, they contribute themselves to the action. These actions give gameplay emotions that define the way a person reacts to a given situation. "It is the player's evaluation of his own coping potential that determine whether the confrontation with a monster will be experienced as fear (if the evaluation of his coping potential is moderate), despair (if he feels that he has no coping potential), or triumphant aggression (if he feels he is amply equipped for the challenge). This entails that the emotional experience will vary over time, because of the learning processes leading to a change in coping potential" [37]. Video games do indeed "simulate emotions in a form that is closer to typical real-life experiences than film," testifies Zillman, "emotions are motivators for actions and are labeled according to the player's active coping potentials" [38].

1) Horror Aesthetics in Video Games: In *Silent Hill* games (Konami 1999–2014), the context is arranged in such a way that keeps the player in a constant state of alert. In addition to the design of the rooms where color of rust and blood dominates, the player is forced to explore the environment in the dark, carrying merely a flashlight to illuminate his surroundings and a radio to warn him against the arrival of monsters. This anxiety behind the lack of visibility, as stated by W. H. Rockett, creates a kind of anticipative, more imaginative, and subtle fear defined as terror. Another aspect of the context that intensifies anxiety is the spread of disturbing and unpredictable noises. Perron states that "this kind of arrhythmic environment sounds resists listener adaptation. The soundscape of the game requires constant attention, which is taxing to the player. This type of low-level affective "trick" can be used to create an unconscious emotional effect that adds fear to the gameplay" [39].

P.T., the playable teaser developed to promote the new game *Silent Hills* (Konami 2014), had already demonstrated how changes in context, whether random or established, play an important role in creating a hostile environment. This game limits the explorable field to an L-shaped corridor that the player is forced to investigate in loop. "Devoid of any real battle preventing a gamer from continuing on, the weird shifts in the environment modify the way in which we perceive the location already visited more than once. This can be done just by having a different reddish illumination or a new writing on the walls" [40].

2) Design of Lighting and Audio in Horror Media:

a) Lighting: A classic perceived effect in horror games is forcing the player to explore the surroundings in the dark. To make it more frustrating than tense, players are never in complete darkness. A *flickering light* from a defective lamp or a dim light from a *flashlight* is always there to guide players but not to reassure them. These effects do not only create a phantasmagorical atmosphere, but do also limit players field of view so that they feel less secure and succumb to the imaginative fear of what may happen [36]. Another very commonly used way in film to manipulate emotions on different levels of consciousness and subconsciousness is through the predominance of a color. Previous work has revealed that the *red color* evokes a negative emotional response associated with a high level of arousal [41].

b) Audio: The horror genre resorts sometimes to misophonic sounds to elicit stress and anxiety. The triggering and constant *ringing of a telephone*, *scream cries of a female*, unpleasant sounds of a *crying baby*, a neutral informative voice of a news anchor on a *radio*, and a heavy *rain* are all sounds registered within the categories defined in the study of Kumar *et al.* [42] on the brain basis for misophonia. Another technique used for the same purpose is forewarning. According to Perron [36], "forewarning intensifies emotional reactions about upcoming frightening events and increases anxiety". One other sound, and maybe the most interesting one to investigate, is the very absence of it. *Silence* is sometimes used in films to arrest the audience's attention over some change in story direction. It may also give at some key point a feeling of emptiness that gradually grows into fear [43].

c) Actions: In games, players are in constant interaction with elements such as windows or doors. Designers, sometimes, take advantage of these interactions to trigger "actions" that would lead players into an intended experience. B. Curtis explained in Dark Places that "in haunted houses, *doors* that refuse to open or close register a fundamental spatial anxiety that portends the sequestration or return of something hitherto feared and partially forgotten" [44]. Another classic technique used to scare people in the survival horror games genre is the trigger of *sudden actions*. This effect manifests through surprise attack, *appearances*, or other disturbing events that happen without warning. However, it is a move considered more and more easy to achieve and therefore "cheap" [36].

III. Game Architecture

The core of the game is built off finite state machines (FSMs) that incorporate some visual and sound effects. Within these FSMs, the events are organized in sequences to create scenarios. The aim is to observe the impact of the events' orchestration on the players' affective state.

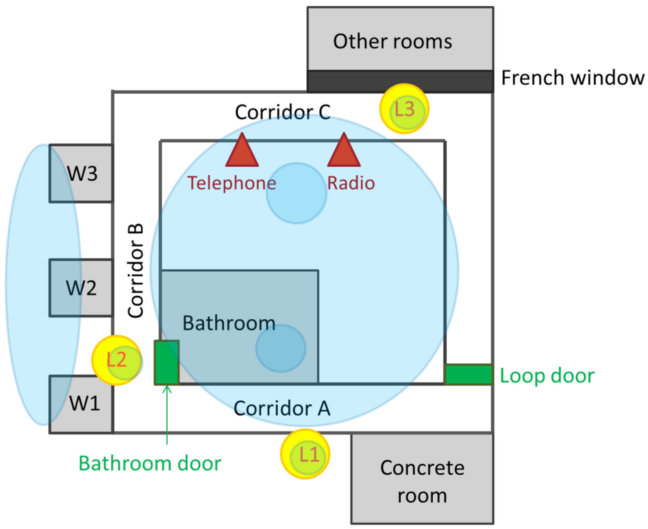

Fig. 1. Game's level architecture. The various components used to manipulate the different effects are illustrated as follows: yellow circles for lamps, green rectangles for doors, red triangles for the telephone and the radio, and blues circles to describe the areas covered by the sound containers.

A. Game: PTVR

PTVR was originally a VR application. This study, though, does not recourse to VR technology as the gameplay experience and player perception of sensory presentation would vary, which could be the subject of a different study. Hence, the game is set up on PC, and players are provided with a mouse, a keyboard to navigate through it, and headphones for a clean audio display. The game is built following a specific logical progression. The real-time changes that occur are meant to set an uneasy climate around the player and to enhance anxiety and confusion, which represent the main studied emotions. The goal of the game is to push forward through the level and escape through the door at the end of the hallway, which will eventually end the "looping effect." In regard of that, the player is forced to go through series of effects embedded in scenarios where each scenario forms a loop within the game. An overview of the gameplay and scenarios is available online.[1]

1) Level Architecture: The explorable space in the game is limited to three areas: a concrete room, a hallway, and a bathroom (see Fig. 1). The concrete room is the starting point of the game and is no longer accessible once the player steps into the hallway. The latter is formed by three corridors shaped in a U. Corridor "A" is the one linked to the loop door and is the first to explore. Corridor "B" is connected to the bathroom and opens up to outside sounds through the windows. Corridor "C" has a sideboard upon which the radio and the telephone are set. Its other side opens up to extra rooms separated by a constantly closed French window. The hallway lights are controlled by lamps placed each in one corridor. Sounds are managed via containers that build-in directional sound, giving a three-dimensional perception of the space. One container was set to report sounds from the radio and the telephone. Three others were associated with each lamp. Two last containers were included: one for outside noises, and the other for the bathroom. An extra one was added to set the general ambiance.

[1][Online Video]: Available: https://youtu.be/WjPWmV77B0g

B. Sounds, Lights, and Actions

For the purposes of this work, the authors explore the manipulation of three diverging "environmental effects" (shortened to effects). Lighting is an important aspect of a virtual scene as it can illuminate or darken certain areas creating different atmospheres. Thus, this work explores its manipulation by interchanging specific effects, which include light intensity, flickering light animation, light color (red), and even the absence of light (darkness). The second consists of audio, which is another important aspect of creating tense atmospheres and environments and include diegetic and nondiegetic sounds. Some are used to add significance to other effects (e.g., click of a flashlight's switch, creaking of a door, etc.), while some others were used to manipulate the overall ambiance and atmosphere. Sounds are also used for narrative purposes and for the construction of tense and "horror"-like atmospheres. Such sounds include a crying baby, screams, a ringing phone, radio alerts, and a heavy rain. It is also important to note that silence (absence of sound) is also used to contrast the frightful sounds and contribute to the tense atmosphere. Finally, the last category referred to as "actions" consists of all "physical" objects that appear or move in the virtual environment such as the appearance of the ghost entity and the movement of doors.

C. Scenarios

A scenario is a sequence of effects triggerable as events during the game. In order to report effectively which effect was efficient at arousing players, a distance must be kept between the times at which events occur. Scenarios were, hence, designed in a way events are most likely not to interfere with each other. To conduct this study, a total of seven scenarios were implemented. This section will give a detailed description of each scenario and highlight the main effects used to help draw the hypotheses on the expected tension intensities of each scenario for the user experiment.

Scenario 1) Radio News: In this scenario, only the radio is playing, reporting news about unusual family murder cases. The place is safe and the player is free to explore the surroundings. This scenario features one main sound effect, which is the radio.

Scenario 2) Heavy Rain: The opening is set by a thunderstrike causing an electric power cut and a quick light restoration. Sounds of falling rain along with radio statics and squeaks of a swaying lamp fill the air. This aims to put the player in a first state of uneasiness. This scenario also exposes one main sound that is the rain.

Scenario 3) Radio Warning: The only remaining sound is that of the radio statics. Before the player reaches the next loop, the radio channel clears to a warning voice announcing that "it is just getting started." This scenario features a warning voice through the radio.

Scenario 4) The Weeping Woman: At the beginning of the scenario, the player is free to walk around. When he is about to reach the next loop, the door is slammed shut in-front of the player with a loud thumping sound. The lights suddenly turn off, and sounds of footsteps and door creaks follow. Eventually, the player is lured to the bathroom by the sound of a weeping woman emanating from that location. After reaching the bathroom door, the player is forced to turn back, with a red light guiding toward corridor "A." Once the player sufficiently walks toward the guided path, the lights suddenly turn on, allowing the player to freely access the next loop. As the player passes by the bathroom, a variety of sounds can be heard on the other side of the door: banging and flesh or squishy-like noises. This scenario features five effects. The woman's cries are the mainly displayed sounds. Two light effects, namely the lights off and the red light, are also included. The two last effects are actions and associates with the main door and the bathroom door.

Scenario 5) Baby Cries and Bathroom Access: Entering this scenario, the player is guided to the bathroom by the sound of a baby cry. At the same time, the lights are continuously flickering. The player is tasked to explore the environment a bit, until the door to the bathroom opens. Once the player enters the bathroom, the door behind slams shut, leaving the player in complete darkness for few seconds. The character eventually turns on a flashlight illuminating the locations the player points toward. After exploring for a few seconds, a red light starts to emanate from under the bathroom door, allowing the player to force the door open again. Forcing the door triggers a loud bang sound, followed by a sinister laugh on the other side. Once the player is out of the bathroom, the rest of the house is dark and silent for the remaining of this scenario. This scenario also features five effects: two sounds relating to the baby crying and silence, two light effects namely the flashlight and the flickering lights, and one action resolving around the bathroom door and accessing the bathroom.

Scenario 6) Rain and Radio Monologue: The place is safe but shrouded in darkness. The only events planned for this scenario are sounds of creaking doors and wood, sounds of heavy rain pouring outside, and a radio monologue directed at the player (or so it seems). This scenario features two sounds: the radio and the rain.

Monster Encounter: Similarly to scenario 4, the door slams shut in front of the player as he tries to reach the next loop. A voice coming from the radio suddenly warns the player alerting him to look behind. Doing this, a blurring effect appears on the player's field of view, and eventually, the monster can be seen for a few seconds on the other side of the French window. Once it disappears, the field of view returns to normal. The player is forced to explore the room some more, until the monster suddenly appears before them for a few seconds. This triggers the phone to ring letting the player know he "has been chosen." The room fills with silence and the player can move on to the next scenario. This scenario features three actions and two sounds. One action relates to the main door and the two remaining are associated with the monster appearance and the blur effect. The sounds refer to a waning voice and the ringing phone.

IV. User Experiment

Visual and sound effects are embedded within the game as events that can either be triggered instantly or occur over time. To investigate the different variations of influence of events orchestration on players' affective state, a user experiment was conducted with human players, with several data types being collected during play.

A. Data Collection Process

For the experiment, four data types are collected. The first type is obtained directly from the game (i.e., game logs), which keeps track of all game events triggered. The second type consists of physiological data (GSR) obtained from the players and synchronized to the in-game events. Real-time emotion annotations are the third type of data collected and consist of self-reported emotional intensity annotations. Finally, postexperimental questionnaires are collected from the players at the end of the playing session.

1) Game's Logging System: The logging system was set to keep track of the instant an event occurs during the game. The moment an event takes place, a timestamp and an identifier along with the event's name are attached to a logging string. A set of parameters specific to each event are eventually included. Synchronization between the logging events and the remaining time-based data (annotations and GSR) are achieved in postprocessing.

2) Physiological Data: For this work, the physiological data collected consist exclusively of GSR. The GSR data stream is monitored by the Empatica E4; a wristwatch-like device worn by the participant during the experiment [45]. The device is controlled via a Bluetooth connection, where all participant data are streamed to the experimental PC. To control the flow of information, synchronization with the game's local timestamp is established.

3) Annotation System: During the experiment, the participant's gameplay is recorded. Right after the playing session has ended, the playthrough is loaded into the RankTrace [46] annotation program allowing the participant to report the intensity of the perceived tension. As the participant annotates, a trail of their previous values over time can be seen, allowing participants to

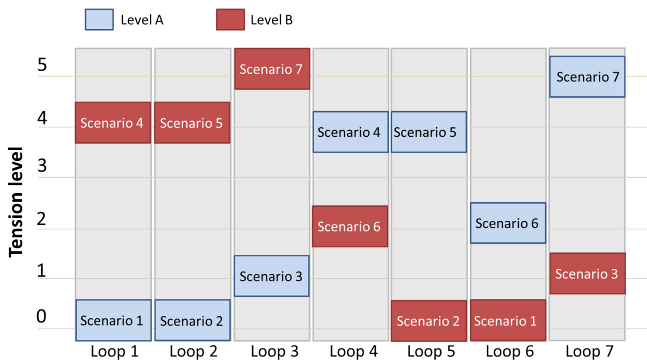

Fig. 2. Expected tension intensity of each scenario.

use it as reference and guide their subsequent annotation. Results of the annotation are synchronized with the game's timeline and are written in an external file.

4) Postexperimental Questionnaire: At the end of the experiment, a questionnaire is presented to the participant. Given recent research on the superiority of rank-based questionnaires [47], a similar methodology was applied in this work. Participants are tasked in ranking the scenarios first and then the different effects, from top to bottom (most to least intense), based on the perceived level of tension.

For easier viewing, each ranked component is distributed across three diverging groups: High (ranked first or second), Low (three bottom lowest ranked), and Medium (ranked in-between the highest and lowest components). The authors made sure of the data's consistency.

B. Levels: Different Scenario Progressions

In order to investigate the individual and the combinatorial affective impact of the events within scenarios, two levels "A" and "B" were designed. In the context of this work, a level consists of a specific progression of different scenarios. The reasoning is that based on the "intended" emotional intensity of each scenario, the way they are ordered can also potentially influence how the player's affective state is induced. The "intended" intensities are hypotheses drawn subjectively by the authors based on the number of the main effects featured in each scenario and their nature (e.g., light, sound, or action). Actions are expected to be the most influential, then sounds, and finally lights. On this basis, level A (see Fig. 2) presents a slow and gradual tension increase over time. It draws scenarios 1–3 first as introduction to the scary atmosphere. Considering the warning effect, scenario 3 was attributed a higher tension intensity than scenarios 1 and 2. Starting from the fourth loop, the scenarios get tenser as the number of the effects increases. Scenario 6 follows and aims to lessen the tension before the big finale. Finally, the level is finished with scenario 7, the most intense expected scenario considering the different actions featured within it. Level B proposes scenarios expected to be the most intense first (scenarios 4, 5, and then 7) and then draws scenario 1–3 and 6 to cut the tension (see Fig. 2). Scenario 3 is set last to avoid that the warning affects the experiment as it is not the focus of the study. These two level progressions form the subject of this study. Other scenario combinations could have been potential options.

C. Experimental Protocol

Initially, the game and controls are presented, and a demo of the annotation program is shown to each participant. Once participants offer their consent to conduct the experiment and finish filling up the demographic survey, the Empatica device is placed upon their wrist. To ensure some form of balance between which version of the game is played first, an alternating method is used where one participant plays A first, while the following plays B first, and so on. All participants are tasked in playing both levels according to the assigned level order. For the purposes of simplicity, version 1 consists of playing level A first and then B, while version 2 consists of the opposite order. After each level, the RankTrace program opens automatically allowing participants to annotate their last playthrough. At the end of the experiment, participants are asked to fill the postexperimental questionnaire ranking the in-game effects based on the experienced tension intensity. Furthermore, it is important to note that the protocol described in addition to the game itself was extensively reviewed and approved by the authors' institution's ethics committee.

D. Feature Extraction: Physiology

The GSR, also known as the electrodermal activity (EDA), is comprised of two main components: the tonic component given by the skin conductance level, which represents the slow-varying baseline level of the skin conductivity, and the phasic component called skin conductance response (SCR), which reflects the evoked response of the eccrine sweat glands to an external stimulus. To extract these features, the Neurokit tool was used [48]. This toolkit uses the cvxEDA method proposed by Greco *et al.* [49], which casts the EDA deconvolution as a quadratic optimization problem. The EDA generation was modeled based on some assumptions. One of the main assumptions states the SCR as a consequence of a burst, generated by a sparse and nonnegative neural signal, from sudomotor nerves controlling the sweat glands [49]. That makes the observation model (y) as follows:

$$y = r + t + \epsilon \quad (1)$$

where r is the phasic, t is the tonic, and ϵ is a noise component

$$r = MA^{-1}p \quad (2)$$

where p represents the sudomotor neuron activity (SMNA) and M and A are tridiagonal matrices. Further details on this method can be found in [49].

This work will focus specifically on SMNA peaks, which will point toward the relevant events that had a significant impact on the players. Thus, Neurokit was applied on the raw GSR signal. The sampling rate was set to 40 Hz, and the minimum threshold for the SCR (in terms of standard deviation) to 0.08 μS [48]. The method renders four types of data: the raw signal, the filtered signal, the phasic component, and a sparse vector corresponding to the SMNA peaks. Along with the logging system data, event extraction consists of the following process.

1) A time window of 5 s before each peak is extracted. The reasoning being that literature estimates that GSR presents a reaction time of 5 s after stimulus and can often take an average of 5 s to drop back down [50].
2) Each time window extracted with the above method may present several in-game events that were triggered simultaneously (or close to simultaneously). Thus, for the purposes of analysis, if this happens, the first event that was triggered within this time window is assumed to be the originator of the peak. However, a special condition is applied, where events must be at least 4 s apart from each other. Otherwise, it is discarded due to the ambiguity of pinpointing the exact originator of the peak in question. Thus, given these conditions, approximately 6.6% of the total data collected was discarded due to inability to associate a peak to a single event.
   One exception was made for two particular events that represent interesting effects to investigate and that could not be dissociated from each other. At this point, the study required the merging of these events into one. These events refer to the blurred vision and the monster behind the window.

For events that occur over time (e.g., audio), the extraction process is different. The assumption is that longer events may induce different reactions and intensities of said reactions from each participant. The event is thus divided evenly into five sections, where each forms a new event labeled same as the main event. Within each section, we observe if at least one peak has occurred. If so, the event is considered relevant.

E. Feature Extraction: Annotations

Annotations can be represented as a signal, where the y-axis is the tension intensity and the x-axis is gameplay time. The features we seek to know are the peaks of tension denoted by participants. To that end, we start by normalizing the signal between [0,1]. Then, we fix consecutive and overlapping windows of 5 s. The windows' length was fixed in consideration of the participants' reaction time at the rise of an event. For each window, we compute the min–max value. A zero value indicates no reaction. Otherwise, a peak is extracted. Decreased intensities suggest participants not being affected by the occurring event, and thus, the event is not taken into account. Once all peaks are extracted, we start identifying the relevant events that induced these peaks similarly as described in the previous section for GSR.

V. Results

Twenty-four participants aged from 18 to 52 took part in the experiment. The average age is approximately 25. Half of the participants consider themselves beginner and average players, while the second half estimate being gamers and expert players. Four people report horror games as their favorite game genre, 11 others state having some experience with these games while the remaining nine have none. Sixty-two percent of the participants (15 participants) admits feeling nervous about playing a horror game.

TABLE I
Minimum, Maximum, Mean, and Standard Deviation Value for the Game Duration (Seconds) Per Level

		μ	σ	min	max
Version 1	**Level A**	475 (s)	49 (s)	405 (s)	570 (s)
	Level B	436 (s)	56 (s)	366 (s)	545 (s)
Version 2	**Level B**	715 (s)	290 (s)	507 (s)	1368 (s)
	Level A	504 (s)	145 (s)	368 (s)	772 (s)

The ANOVA test is calculated pairwisely for the first and last played levels in each version. Levels marked in bold are significant ($p < 0.05$).

TABLE II
Average Duration (Seconds) of Each Scenario per Level and Version Combination

	μ (s) Version 1		μ (s) Version 2	
	Level A	Level B	Level B	Level A
Scenario 1	55.416 (s)	48.853 (s)	57.417 (s)	49.854 (s)
Scenario 2	32.429 (s)	34.722 (s)	51.816 (s)	39.109 (s)
Scenario 3	30.710 (s)	17.181 (s)	25.635 (s)	26.641 (s)
Scenario 4	82.006 (s)	72.182 (s)	124.641 (s)	75.111 (s)
Scenario 5	151.965 (s)	114.703 (s)	249.619 (s)	178.409 (s)
Scenario 6	20.057 (s)	34.618 (s)	50.34 (s)	39.793 (s)
Scenario 7	58.980 (s)	63.672 (s)	108.800 (s)	53.351 (s)

Significance is calculated using ANOVA by first combining all scenario playthroughs and comparing it pairwisely among each other. Scenarios marked in bold present significant differences ($p < 0.05$) with all other scenarios.

A. Playtime Analysis

All participants played both levels, and no time restriction was imposed on the experimental process. Out of the 48 one-level runs, only 32 were selected due to unusable data captured by the device. Overall, eight runs performed by participants with mixed skills were recorded for each level making it a total of 16 runs each version. It is important to keep in mind that players' skill was not taken into consideration in the analysis. In fact, the game itself consists in moving around an environment and thus does not require much skill. It is one of the reasons why a game such as *P.T.* was interesting to explore as players are not required to perform dexterous or any extensive gameplaying maneuvers.

Table I presents a statistical analysis of the game duration per level. On average, participants spent less time on the second played levels (level B for version 1 and level A for version 2). According to Table II, participants playing level B first spent almost twice the amount of time of A, even though both presented similar content. The difference is prominent for scenarios 4, 5, and 7. In fact, both scenarios 4 and 5 show strong statistical significance ($p < 0.01$) when paired with all remaining scenarios. Scenario 7, though, is significant when paired with scenarios 3 and 6 only.

B. Results of the Postexperimental Questionnaire

Fig. 3 showcases results for ranking scenarios. Scenario 4 presents a consistently high-ranking scenario among most participants, suggesting that this scenario was the most effective in creating a tense atmosphere. Scenario 5 is proposed second most

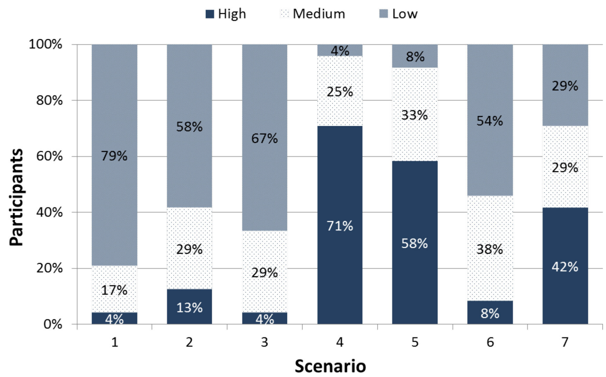

Fig. 3. Scenarios distributed by high, average, and low ranks based on the participant's postexperimental ranking questionnaire (High: rank 1 and 2; Average: rank 3 and 4; and Low: ranks 5–7).

TABLE III
RESULT OF THE KRUSKAL–WALLIS H TEST FOR RANKED SCENARIOS

Scenarios	1	2	3	4	5	6	7
1	–	–	–	–	–	–	–
2	2.506	–	–	–	–	–	–
3	0.842	0.559	–	–	–	–	–
4	29.943**	20.657**	26.920**	–	–	–	–
5	25.390**	15.678**	21.756**	0.869	–	–	–
6	3.178	0.016	0.848	22.153**	16.735**	–	–
7	13.399**	5.740	9.694**	5.489	2.414	5.876*	–

This test's p-values were corrected using the Bonferroni method. Statistical significance is represented by (*) for $p < 0.05$ and (**) for $p < 0.01$.

influencing scenario as more than half of the participants rank it as high and only a minority as low. Scenario 7, in contrast with the previous hypotheses (see Fig. 2), is ranked third. These three scenarios are the only which observed statistical differences (see Table III). Finally, scenarios 2 and 6, which feature the radio and rain sounds together, were ranked low average, making them slightly more influencing than in the hypotheses.

Looking exclusively at the type of effect used (i.e., sound, lights, or actions) suggests that sound tended to have a higher impact than both other effects. In fact, participants tended to consistently favor sound over the others. Actions presented the most division among participants, suggesting some ambiguity on its effectiveness overall.

The rest of the questionnaire details lights, sounds, and actions according to their respective effects. Results on light rankings suggest that the lights OFF effect is the most striking effect. The latter accumulated a total of 71% of high ranks and had only three participants stating it as low. The flickering lights event hold second place with a record for high ranks on over half of the experiments. Among the rest of participants, 29% stated it as average and 17% as low. As for the red light effect, even though ranked high by the majority of the participants, a third of them ranked it as low. The flashlight event seems to bring some sort of ambiguity as the collected rankings are evenly distributed over the three given classes.

Sound of the radio was ranked low by half of the participants. Thirty-five percent considered it average, and only 13% rated it high. Although in the previous analysis, scenarios highlighting the sound of the rain were slightly distinguished (scenarios 2 and 6), this sound perceived individually seems to be the least influencing. It was ranked low by more than half of the participants and average by 29% of them. The woman's scream cry sound is the most striking effect and recorded 75% of high ranks. Only two participants stated it as low. The cries of the baby represent the second most influencing sound with a total of 71% of high ranks and 17% of average ranks. Silence is medium-high with a total of 33% for high ranks and 38% for average ranks. Results for Action suggest actions around doors as the most effective in creating tension. The bathroom door event presents a consequent high ranking among the majority of participants with only 17% of low estimations, while the main door event showcase 46% of high ranks and 21% of average ranks. The blurred vision effect is the least influencing action with 63% of low ranks. The monster encounter had the most conflicting results among them, presenting a clear division between high and low intensities with 42% and 50%, respectively.

C. Stimulus-Reaction Distance Between GSR and Annotation

To study the impact of different faceted game events, two types of data were collected: physiological reactions through GSR and real-time user annotations of tension (RankTrace). This analysis attempts to distinguish the differences between both data types by comparing the average *relevance rates* of each level. To obtain a relevance rate, it is first necessary to extract the detected peaks of both physiological data, sudomotor nerve activity peaks (SMNA), and of their annotations. This provides an objective measure of in-game events that had some degree of impact on the player. The relevance rates are then computed using these peaks, for both GSR and annotations, as follows:

$$r\text{GSR}_i = \frac{n\text{GSR}_i}{t\text{Max}_i}, \qquad r\text{Annot}_i = \frac{n\text{Annot}_i}{t\text{Max}_i} \tag{3}$$

where $r\text{GSR}_i$ and $r\text{Annot}_i$ are the respective relevance rates of a given event i for GSR and annotations, and $n\text{GSR}_i$ and $n\text{Annot}_i$ represent the respective number of times participants reacted to one specific event i physiologically and through annotations. $t\text{Max}_i$ is the number of maximum triggers of one event i during a level.

Given the spatial constraints of this article, the statistical results reported in this article were condensed to the most relevant findings obtained. Furthermore, given that the levels that were played first (Level A for version 1 and Level B for version 2) had a higher statistical impact, this analysis will focus specifically on these. Thus, from this point forward when discussing Level A or Level B, it will be the version in which it was played first, unless it is stated otherwise. Fig. 4 presents the relevant rates obtained for each of the events developed for the specific facets being explored: light, action, and sound.

1) Lights: This section takes a closer look at the light events specifically, which include the flashlight, light flicker, lights OFF, and red light events [see Fig. 4(a)].

Some interesting differences are observed for the majority of the light events with the exception of the red light, which presented a small difference between both versions of the game. Furthermore, level A tended to present a consistent minor error between the GSR and annotation values; however, this does not necessarily suggest that there is a correlation between both (see

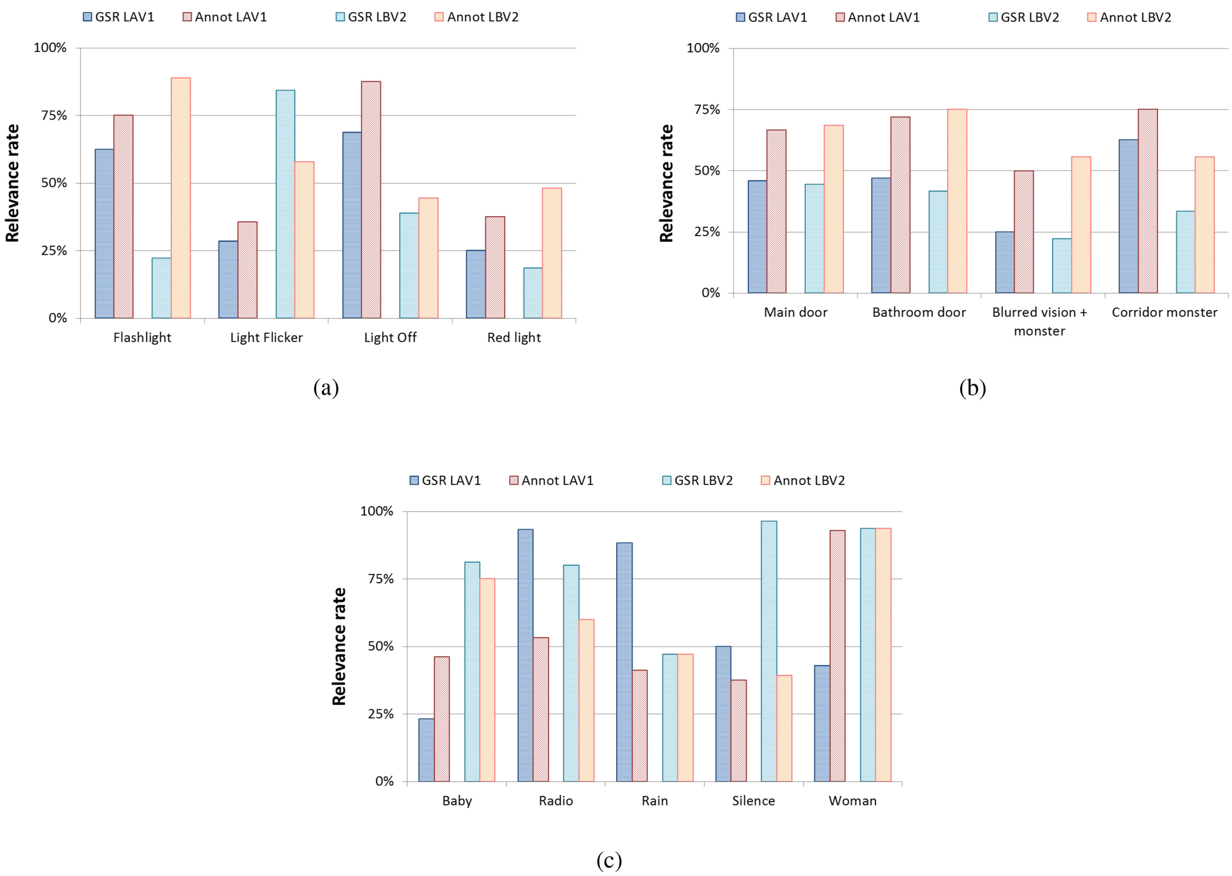

Fig. 4. GSR and annotation (Annot) relevance rate calculation of the first played levels, Level A Version 1 (LAV1) and Level B Version 2 (LBV2), for each event pertaining to the specific facets explored in this article (light, action, and sound). (a) GSR and annotation relevance rate for *Light Events*. (b) GSR and annotation relevance rate for *Action Events*. (c) GSR and annotation relevance rate for *Sound Events*.

Section V-D), but that during play and the annotation phase, the participants tended to consistently react to this event when it occurred. The flashlight and turning OFF the light events tended to obtain the most consistent reactions from participants in both data types collected.

In contrast, for Level B, some discrepancies can be observed. For the flashlight event, a large difference exists between GSR and annotation, suggesting that participants tended to consistently perceive this section as tense, even though physiological reactions were rarely encountered. This may be due to the fact that the flashlight section tended to be close (timewise) to the "crying woman" sound event, which consistently obtained a high relevance rate and may have influenced the overall perception. Contrarily, the light flicker presented an opposite phenomenon, where participants tended to underestimate the perceived tension in relation to the physiological reactions. Though, this difference between GSR and annotation data is not as high as the previous comparison.

2) Actions: Fig. 4(b) presents the relevance rate of all events designed for the action facet, which include main door (opening and closing), bathroom door (opening and closing), apparition of the monster with a blurry effect, and apparition of the monster at the end of the corridor.

As a first observation, it can be seen that a similar pattern emerges across the different actions events, where participants tend to consistently perceive an emotional reaction more often during the annotation when compared to physiology. Interestingly, participants tended to have a higher physiological reaction for the monster reveal event of Level A when compared to Level B. This suggests that the distribution of scenarios for Level A, which followed the natural "suspense arc," did indeed provide a more effective way of evoking a reaction than the arc defined in Level B. This pattern also emerges in the annotation, suggesting that the participants also perceived it this way.

Apart from the latter, the consistency among the other action events suggests that the scenario distribution was not particularly influential on these events.

3) Sounds: This section analyzes all events specifically related to sound, which include baby crying, radio chatter, rain, silence (absence of sound), and woman crying [see Fig. 4(c)].

Of all the facets explored within this work, sounds obtained the highest rates when compared to the other two facets. Overall participants tended to physiologically react to most of the audio cues, although it can vary substantially based on the level variant. For example, the baby cry audio cue had a higher relevance rate in the Level B playthrough rather than Level A. This may be due to the fact that this event appears right at the beginning of Level B, which may catch some participants off-guard.

The radio, in particular, also obtained a strong GSR relevance rate across both level variants. It is important to remember that

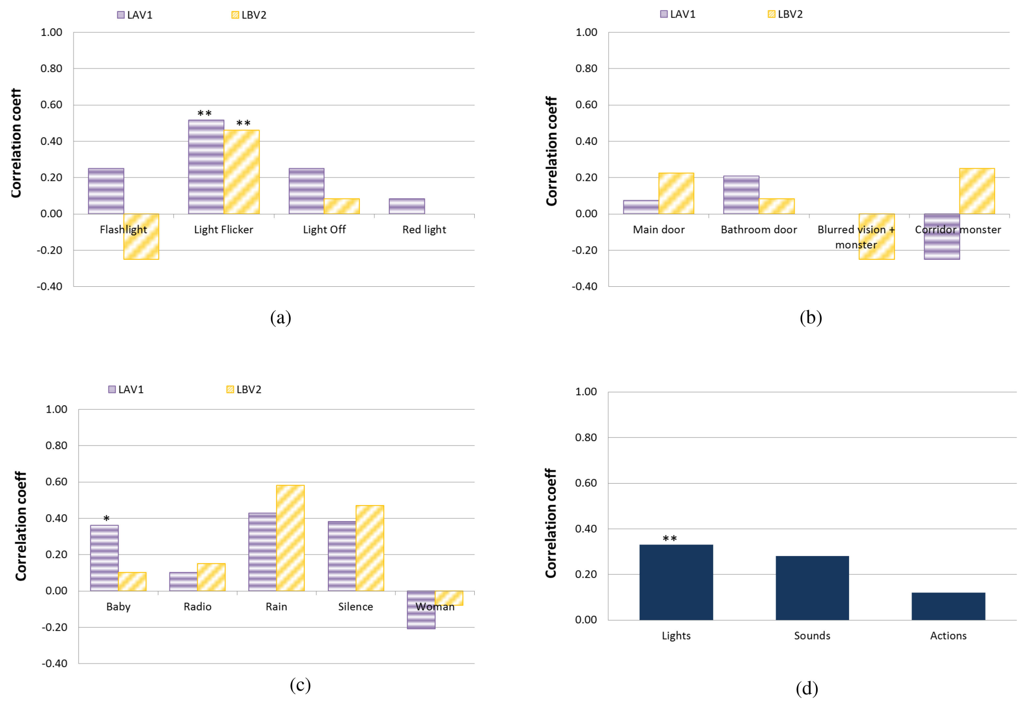

Fig. 5. Binomially distributed pairwise correlation coefficient between GSR and annotation of Level A (LAV1) and Level B (LBV2). Significance is calculated using the Kendall τ significance test, where $p < 0.05$ is represented by (*) and $p < 0.01$ by (**). (a) Correlation of each *light event*. (b) Correlation of each *action event*. (c) Correlation of each *sound event*. (d) Correlation of LAV1 and LBV2 of all events in each facet.

this particular event provided a combination of both narrative and audio facets and as such was not a "raw" sound but human dialogue. Thus, this contextualization of the players plight may have increased the efficiency in obtaining physiological reactions from the player. It may also explain why Level A obtained a higher relevance rate to Level B, given the narrative arc presented by the "radio host." Even though the narrative is ambiguous enough to be presented in any order, the original arc is in fact the one presented in Level A.

Physiological reactions were also consistently present during the rain sequences in Level A; however, the discrepancy between physiology and annotation may be similar to the phenomenon observed for silence in Level B, as these sequences are often presented after an intense emotional stimulus. As for Level B, the tension starts immediately high with a strong stimulus straight at the start of the game (i.e., scenario 5), which may lead to a stronger reaction to the silent section that immediately follows, while for Level A, scenario 6 is similarly followed by scenario 5, which may result in a more tense experience during this section. Overall, these results may be due to the fact that players are "on their guard" meaning that even though the scenario presents little to no stimulus (tensionwise), the effects from the previous event may still be on their minds, thus also explaining the discrepancy between physiology and annotation on both situations.

Finally, the sound of the weeping woman presented the most consistent reaction from annotators for both levels. For Level B, the player annotations and physiological reactions followed a similar trend, meaning that this scenario did in fact impact the players in some way. Furthermore, given that this scenario was the first one presented to the players of Level B may explain why its impact was more apparent than that of Level A. The reported perceived tension of Level A similarly matches that of Level B, meaning that perceptually this scenario is considered tense for both levels; however, in terms of physiological reactions, the findings did not coincide.

D. Binomially Distributed Pairwise Correlation of GSR and Annotation

This method computes the correlation between GSR and annotation data using the binomially distributed pairwise methodology presented in [51]. It computes $c(z) = (1/N)\sum_{i=1}^{N} z_i$, where N is the number of samples. To achieve this, GSR and annotation data are set in pairs for each experiment. Then, peaks for each signal are extracted. We consider afterward windows of 5 s around events within which we observe if a peak has occurred. Thus, if a peak is spotted during the same time interval for both GSR and annotation, $z_i = +1$. In addition, if no peak is observed, $z_i = +1$. Else, $z_i = -1$. Once all N events are processed, the correlation coefficient $c(z)$ is computed. Furthermore, to calculate the significance of each correlation, the Kendall τ significance test was used. Fig. 5 showcases the correlation values obtained using the aforementioned methodologies.

1) Lights: Out of all the different facets explored within the context of this work, lights presented the only strong significant results [see Fig. 5(a)]. Specifically, the light flicker events had a very strong correlation (i.e., $p < 0.01$) of around ≈ 0.52 and 0.55 for Levels B and A, respectively. Given that this event occurred in situations of high tension and after the baby crying sequence (which also presents a significant correlation for Level A) suggests that this sequence of events may have had a significant impact on the players. Other events did not present such a strong correlation, where the lights OFF event had minor correlations for both level variants, while the flashlight event presented a positive and negative minor correlations for Levels A and B, respectively. This difference may be explained given the previous observed results in Level B [see Fig. 4(a)], where a large disparity between the relevance rates of annotations and GSR is seen, suggesting that annotators perceived the event as very tense even though there was not a lot of physiological reactions observed.

2) Actions: The action facet presented results without any definitive conclusions [see Fig. 5(b)]. The most interesting results were on the monster reveal, which presented a negative and positive minor correlation for Levels A and B, respectively. Although the correlation is not significant, it does suggest a dissonance between the annotation and GSR, even though the relevance rates were quite high and similar for both data types and level variants.

3) Sounds: Fig. 5(c) showcases the correlations obtained for each sound event explored within this work. Most events presented positive correlations for both level variants with the exception of the "woman weeping" sound event. Although not significant, the latter result is interesting, given that the previous observation of this same event presented the highest relevance rate out of all other sound events. Given that this event occurs over the course of several seconds may be the cause of this discrepancy where a time window dissonance between the extracted peaks of annotation and physiology occurs. In contrast, the only significant correlation observed was the "baby" event for the Level A variant but not Level B. The rain and silence events obtained the highest correlations out of this group, albeit not significant. Finally, the radio event presented minor positive correlations for both levels, even though it consistently obtained high relevance rates for either of them. Similarly to the "weeping woman" event, there may be a dissonance between both data types during the extraction process. Furthermore, this event is one of the few that reoccur over the course of the game (i.e., in multiple scenarios), which may further enhance the aforementioned phenomena.

4) Combined: The last analysis consists of observing the combined event correlation of both Levels A and B for each individual facet explored within this work [see Fig. 5(d)]. The only significant correlation observed is related to the light facets, which was expected considering the previous correlations observed, which were the only ones that had obtained a significant result. Although sounds also obtained a moderate correlation similarly to lights, no significance was observed. Given the previous results, it can be assumed that certain events influenced the lower correlation and the lack of significance, in particular the negative correlations for the "weeping woman" event and the correlation discrepancy between the "baby" event in Levels A and B. Finally, the actions facet presented a minor nonsignificant correlation, which was also expected given the mixed correlation results analyzed previously.

VI. Discussion

Playtime results show that, on average, participants spent more time going through level B than level A. This may be due to several factors, which may not be directly related to the "tension factor." For example, the first scenarios presented in Level B were more complex than those of Level A, which given the lack of previous experience to the game may have provided a more confusing and difficult start when compared to Level A. Furthermore, for the latter, the initial scenarios are straightforward lacking any "complex" puzzle elements, which allowed players to understand the concept of the game and the objective immediately explaining why Level A players spent less time in each scenario. However, this confusion for the Level B players potentially enhanced the overall tension, given that relevance rate observed for the events contained in initial scenarios of Level B tended to be much higher for these players than those of Level A. Thus, this lack of direction and objective did contribute to the overall experience.

Although the majority of effects did tend to stimulate participants perceptually (through annotation) and physiologically (through GSR), the sound facet events tended to achieve the highest relevance rates for both data types. This can also be observed in the scenario rankings, where participants tended to rank the scenario with the weeping woman (scenario 4) and baby (scenario 5) as the most tense, respectively. This also goes in line with the findings in the other facets, where the highest relevant events tended to be the ones that are present within these two scenarios such as the light flicker (scenario 5), silence (scenario 4), lights off (scenarios 4 and 5). Scenario 6 is also interesting given that it was the scenario that obtained a mixed ranking with no clear result in-between mid to low tension, which can also be partially observed for the rain event relevance rate for Level A, which was quite high for GSR but not as much when examining the annotation. This is a situation where the sequence of scenarios may have played a part in the overall perception of this scenario, whereas for Level A, this scenario comes immediately after scenarios 4 and 5, which may have provoked players to be "on edge." In contrast, for Level B, this scenario comes immediately after the monster reveal, which might have allowed time for players to calm down as the danger "had passed."

Analyzing the correlation between GSR and annotation obtained the most inconclusive results, which may be due to the nature of how the data were extracted. Although for the light flicker event, we observed highly significant correlations ($p < 0.01$) for both Levels A and B and an additional significant correlation ($p < 0.05$) for the baby event exclusively in Level A, no others were observed. The significant consistency observed for the light flicker event may derive from the overall scenario, which occurred over the course of several other events such as the baby event, which may explain the significance observed in Level A. However, correlation between GSR and annotation only represents a "concordance" measure between both the perceived tension and physiological reactions; thus, a high significant correlation does not necessarily mean that the event was actually effective in stimulating the intended emotion. However, it does suggest that some emotional impact was perceived and felt by the player simultaneously.

VII. Conclusion

Gaming experiences represent the keystone behind video games' building process. The empathy toward ones character and the fear over its life, the accomplishment behind puzzles' solving, and the challenge to surpass oneself are all experiences sought by players. In horror games, the orchestration of audio-visual effects plays a key role in shaping the fearful atmosphere. Understanding which effects to embed is thus an important step to build the intended gaming experience. To investigate how orchestration of visuals, audio, and in-game interactions can be used to induce negative emotions, *P.T.* environment was chosen in regard of its fixed level architecture. An experiment was conducted, and both physiological and self-reported data were collected. Two methods were developed to study the correlation between these two classes of data. Results suggest the sound facet events as the most effective in stimulating stress/anxiety/tension as they tended to achieve the highest relevance rates for both data types.

In the horror genre, scenarios usually follow a well-established sequence of events that gradually introduce the audience to a well-targeted scary atmosphere. And since fear is the aimed emotion, it provides a potential measure of tension and stress. These factors set a favorable ground to conduct our study. Although results might be strongly biased on this genre in particular, it does not reject the possibility of applying these techniques with other game genres/categories, since both visual and sound effects are important blocs to build in all games' context.

Due to the limited number of experiments, hypotheses about scenarios' tension level were drawn subjectively, and only two level progressions were investigated. A more rigorous approach could have been to assign scenarios to levels randomly and study all different combination. This could have allowed a better insight regarding the individual and combinatorial affective impact of each effect.

The study of sounds in this project focused mainly on the diegetic sounds ignoring the impact of those used to add to the ambiance. The aim was in fact to understand players' physiological reactions and the level of awareness of their own emotions toward certain audiovisual variations and not to focus on one specific effect. However, the fusion between these different diegetic and nondiegetic sounds with the rest of the visual effects might give rise to divergent affective responses. It would, therefore, be interesting to investigate the individual impact of these nondiegetic sounds with different parameter settings such as loudness and pitch [52] and observe the result of their orchestration with other effects.

In the future, we want to apply this knowledge for autonomous systems capable of adjusting these environmental variables according to human emotion. Hence, scenarios can be generated in real time while adapting to the player's responses, based on GSR data. How accurate can the decisions made be, based solely on GSR data? What would be the impact if we were to use PCG algorithms to create new audiovisual content similar to some relevant effects How to identify similar effects?

Acknowledgment

The authors would like to thank the participants in the experiments.

References

[1] G. N. Yannakakis and J. Togelius, "Experience-driven procedural content generation," *IEEE Trans. Affect. Comput.*, vol. 2, no. 3, pp. 147–161, Jul.–Sep. 2011.

[2] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Pers. Social Psychol.*, vol. 17, no. 2, pp. 124–129, 1971.

[3] S. S. Tomkins, "Illuminating and stimulating. (Book reviews: Affect, imagery, consciousness. vol. 1, the positive affects)," *Sci.*, vol. 139, pp. 400–401, 1963.

[4] L. F. Barrett and J. A. Russell, "Independence and bipolarity in the structure of current affect," *J. Pers. Social Psychol.*, vol. 74, no. 4, pp. 967–984, 1998.

[5] R. W. Picard, "Affective computing: Challenges," *Int. J. Human-Comput. Stud.*, vol. 59, nos. 1/2, pp. 55–64, 2003.

[6] J. T. Cacioppo *et al.*, "The psychophysiology of emotion," in *Handbook of Emotions*, vol. 2. New York, NY, USA: Guilford Press, 2000, pp. 173–191.

[7] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cogn. Emotion*, vol. 23, no. 2, pp. 209–237, 2009.

[8] S. Balters and M. Steinert, "Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices," *J. Intell. Manuf.*, vol. 28, no. 7, pp. 1585–1607, 2017.

[9] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychol.*, vol. 14, no. 4, pp. 261–292, 1996.

[10] M. D. Robinson and G. L. Clore, "Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes," *J. Pers. Soc. Psychol.*, vol. 83, no. 1, pp. 198–215, 2002.

[11] A. Jaimes, T. Nagamine, J. Liu, K. Omura, and N. Sebe, "Affective meeting video analysis," in *Proc. ICME IEEE Int. Conf. Multimedia Expo.*, 2005, pp. 1412–1415.

[12] C. Tsigos, I. Kyrou, E. Kassi, and G. Chrousos, *Stress, Endocrine Physiology and Pathophysiology*. South Dartmouth, MA, USA: MDText.com, Inc., 2016.

[13] W. D. Fenz and S. Epstein, "Gradients of physiological arousal in parachutists as a function of an approaching jump," *Psychosomatic Med.*, vol. 29, no. 1, pp. 33–51, 1967.

[14] R. McCraty, M. Atkinson, W. A. Tiller, G. Rein, and A. D. Watkins, "The effects of emotions on short-term power spectrum analysis of heart rate variability," *Amer. J. Cardiol.*, vol. 76, no. 14, pp. 1089–1093, 1995.

[15] J. Sugenoya, S. Iwase, T. Mano, and T. Ogawa, "Identification of sudomotor activity in cutaneous sympathetic nerves using sweat expulsion as the effector response," *Eur. J. Appl. Physiol. Occupat. Physiol.*, vol. 61, nos. 3/4, pp. 302–308, 1990.

[16] H. D. Critchley, R. Elliott, C. J. Mathias, and R. J. Dolan, "Neural activity relating to generation and representation of galvanic skin conductance responses: A functional magnetic resonance imaging study," *J. Neurosci.*, vol. 20, no. 8, pp. 3033–3040, 2000.

[17] D. Civitello, D. Finn, M. Flood, E. Salievski, M. Schwarz, and Z. Storck, "How do physiological responses such as respiratory frequency, heart rate, and galvanic skin response (GSR) change under emotional stress?" *J. Adv. Stud. Sci.*, vol. 1, pp. 1–20, 2014.

[18] N. Craft and J. B. Schwartz, "Effects of age on intrinsic heart rate, heart rate variability, and AV conduction in healthy humans," *Amer. J. Physiol.-Heart Circul. Physi.*, vol. 268, no. 4, pp. H1441–H1452, 1995.

[19] D. Mendelowitz, "Advances in parasympathetic control of heart rate and cardiac function," *Physi.*, vol. 14, no. 4, pp. 155–161, 1999.

[20] N. Depraz and A. J. Steinbock, *Surprise: An Emotion?* New York, NY, USA: Springer, 2018.

[21] K. Isbister, *How Games Move Us: Emotion by Design*. Cambridge, MA, USA: MIT Press, 2016.

[22] G. N. Yannakakis and A. Paiva, "Emotion in games," in *The Oxford Handbook of Affective Computing*. Oxford, U.K.: Oxford Univ. Press, 2014, pp. 459–471.

[23] M. Csikszentmihalyi, S. Abuhamdeh, and J. Nakamura, "Flow," in *Flow and the Foundations of Positive Psychology*. New York, NY, USA: Springer, 2014, pp. 227–238.

[24] B. Cowley, D. Charles, M. Black, and R. Hickey, "Toward an understanding of flow in video games," *Comput. Entertainment*, vol. 6, no. 2, 2008, Art. no. 20.

[25] A. Liapis, G. N. Yannakakis, and J. Togelius, "Computational game creativity," in *Proc. 5th Int. Conf. Comput. Creativity*, 2014, pp. 46–53.

[26] A. K. Hoover, W. Cachia, A. Liapis, and G. N. Yannakakis, "Audioinspace: Exploring the creative fusion of generative audio, visuals and gameplay," in *Proc. Int. Conf. Evol. Biol. Inspired Music Art*, 2015, pp. 101–112.

[27] M. Cook and S. Colton, "Multi-faceted evolution of simple arcade games," in *Proc. IEEE Conf. Comput. Intell. Games*, 2011, pp. 289–296.
[28] P. Lopes, A. Liapis, and G. N. Yannakakis, "Sonancia: Sonification of procedurally generated game levels," in *Proc. 1st Comput. Creativity Games Workshop*, 2015.
[29] M. A. Palmer, "Fear: A psychophysiological study of horror film viewing," Unpublished thesis, Texas State Univ., San Marcos, TX, USA, 2008.
[30] C. McCauley, "When screen violence is not attractive," in *Why We Watch: The Attractions of Violent Entertainment*. New York, NY, USA: Oxford Univ. Press, 1998, pp. 144–162.
[31] N. Carroll, *et al.*, *The Philosophy of Horror: Or, Paradoxes of the Heart*. Abingdon, U.K.: Routledge, 2003.
[32] M. Zuckerman, *Sensation Seeking*. Hoboken, NJ, USA: Wiley, 1979.
[33] C. Rickey, "Hooked on horror: Why we like scary movies," *Mademoiselle*, vol. 169, pp. 168–170, 1982.
[34] S. King, *Danse Macabre*. New York, NY, USA: Simon and Schuster, 2010.
[35] B. F. Skinner, *Contingencies of Reinforcement: A Theoretical Analysis*, vol. 3. Cambridge, MA, USA: BF Skinner Foundation, 2014.
[36] B. Perron, "Sign of a threat: The effects of warning systems in survival horror games," in *Proc. COSIGN Conf.*, 2004, pp. 132–141.
[37] T. Grodal, "Stories for eye, ear, and muscles," in *The Video Game Theory Reader*. Abingdon, U.K.: Routledge, 2003, pp. 129–155.
[38] D. Zillmann and P. Vorderer, *Media Entertainment: The Psychology of Its Appeal*. Abingdon, U.K.: Routledge, 2000.
[39] B. Perron, *Horror Video Games: Essays on the Fusion of Fear and Play*. Jefferson, NC, USA: McFarland, 2009.
[40] B. Perron, *The World of Scary Video Games: A Study in Videoludic Horror*. New York, NY, USA: Bloomsbury, 2018.
[41] E. Joosten, G. Van Lankveld, and P. Spronck, "Influencing player emotions using colors," *J. Intell. Comput.*, vol. 3, no. 2, pp. 76–86, 2012.
[42] S. Kumar *et al.*, "The brain basis for misophonia," *Current Biol.*, vol. 27, no. 4, pp. 527–533, 2017.
[43] T. Takemitsu, Y. Kakudo, G. Glasow, and S. Ozawa, *Confronting Silence: Selected Writings*. New York, NY, USA: Scarecrow, 1995, vol. 1.
[44] B. Perron, "Coming to play at frightening yourself: Welcome to the world of horror video games," in *Proc. Aesthetics Play Conf.*, 2005, pp. 14–15.
[45] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti, "Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition," in *Proc. IEEE EAI 4th Int. Conf. Wireless Mobile Commun. Healthcare*, 2014, pp. 39–42.
[46] P. Lopes, G. N. Yannakakis, and A. Liapis, "Ranktrace: Relative and unbounded affect annotation," in *Proc. IEEE 7th Int. Conf. Affect. Comput. Intell. Interact.*, 2017, pp. 158–163.
[47] H. Martinez, G. Yannakakis, and J. Hallam, "Don't classify ratings of affect; rank them!" *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 314–326, Jul.–Sep. 2014.
[48] D. Makowski, "Neurokit" 2016. [Online]. Available: https://github.com/neuropsychology/neurokit.py
[49] A. Greco, G. Valenza, and E. P. Scilingo, *Advances in Electrodermal Activity Processing with Applications for Mental Health*. New York, NY, USA: Springer, 2016.
[50] iMotions, "GSR pocket guide," 2016. [Online]. Available: http://www.imotions.com/about
[51] G. N. Yannakakis and J. Hallam, "Ranking vs. preference: A comparative study of self-reporting," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, 2011, pp. 437–446.
[52] T. Garner, M. Grimshaw, and D. A. Nabi, "A preliminary experiment to assess the fear value of preselected sound parameters in a survival horror game," in *Proc. 5th Audio Mostly Conf.: A Conf. Interact. Sound*, 2010, pp. 1–9.

Sarra Graja is working toward the Postgraduate degree at Abertay University, Dundee, U.K.

In 2018, she was with the Social Intelligence and Multi-Sensing Group and the Computer Vision and Multimedia Laboratory, University of Geneva, Geneva, Switzerland, where she developed an interest for affective computing. Her current research interests include learning to recognize and interpret human affect through physiological measures in order to build affect-aware games and enhance gaming experience.

Phil Lopes (Member, IEEE) received the Ph.D. degree in artificial intelligence in digital games from the University of Malta, Msida, Malta, in 2017.

He is currently a Postdoctoral Researcher with the Immersive Interaction Group, École Polytechnique fédérale de Lausanne, Lausanne, Switzerland. He explored concepts at the interplay of affective computing, game orchestration, and procedural content generation with the University of Malta. His current research interests include developing tools within the domains of audio, level design, and visuals to explore its capabilities of inducing human emotion, while also exploring new ways of orchestrating this type of content so as to provide personalized and unique experience for players. Furthermore, he is also exploring ideas such as game adaptation and its applicability in different domains of digital games and contexts.

Guillaume Chanel received the Ph.D. degree in computer science from the University of Geneva, Geneva, Switzerland, in 2009.

He worked on machine learning for the automatic assessment of emotions based on EEG and peripheral signals with the University of Geneva. From 2009 to 2010, he was at the KML-Knowledge Media Laboratory, Aalto University, Helsinki, Finland, studying the physiological correlates of social processes taking place between players during video gaming. He is currently the Head of the Social Intelligence and Multi-Sensing Group and a Senior Lecturer jointly affiliated with the Swiss Center for Affective Sciences and with the Department of Computer Science, University of Geneva. His research investigates how machines can learn to behave in a social and affective environment. He is particularly interested in the use of multimodal and physiological measures for entertainment and for improving man–machine and human remote interactions. Examples of his research include: dynamic adjustment of games mechanic based on players' emotions, inclusion of physiological emotional cues in mediated social interactions, movie highlight detection based on spectators' social reactions, and adaptation of human social behaviors through machines.