Classification of Affective Data to Evaluate the Level Design in a Role-playing Videogame

Fabrizio Balducci, Costantino Grana, Rita Cucchiara

Dipartimento di Ingegneria "Enzo Ferrari"
Università degli Studi di Modena e Reggio Emilia
Via Vivarelli 10, Modena (MO) 41125, Italy
Email: {name.surname}@unimore.it

Abstract—**This paper presents a novel approach to evaluate game level design strategies, applied to role playing games. Following a set of well defined guidelines, two game levels were designed for Neverwinter Nights 2 to manipulate particular emotions like boredom or flow, and tested by 13 subjects wearing a brain computer interface helmet. A set of features was extracted from the affective data logs and used to classify different parts of the gaming sessions, to verify the correspondence of the original level aims and the effective results on people emotions. The very interesting correlations observed, suggest that the technique is extensible to other similar evaluation tasks.**

I. Introduction

The concept of playing is strictly linked to our lives as one of the first activities that a human experiences, discovering his surrounding world: one method to learn rules or how to safely execute a task is to see it like a gaming activity, gaining experience for future goals [1].

Frasca [2], referring to Csíkszentmihályi's studies [3], defines the modern concept of *ludology* like the "science of the game" which uses research methods and theories from a wide range of scientific communities such as HCI and psychophysiology, with the aim to improve the methodologies to study both players and (video)games: this should help to understand the processes involved in the design of an optimal *player experience*, i.e. the set of feelings and opinions that come from an activity which has to capture the player's senses offering amusement and engagement.

In recent decades the videogame has become a modern entertainment medium with expensive budget, which can rival with cinema, music, comics and literature; considering that every person is different because of factors like age, education, country and personal history it becomes crucial attempting to define and manipulate complex concepts like "the fun" considering also the different player's feelings and preferences. Nacke [4] consider the *affective ludology* referring to investigations of affective player-game interaction to understand emotional and cognitive player experiences: it should investigate cognition, emotion, and goal-oriented behavior of players from a scientific perspective by establishing rigorous methodologies (e.g. psychological player testing or physiological response analysis of players).

Mandryk and Inkpen [5] have successfully demonstrated how electroencephalography can evidence human emotions and cognitive activity during gaming. A brain-computer interface (BCI) is a communication systems that measures brain activity with sensors positioned on the subject's scalp: the brain is a dense network of neurons, nerves cells which communicate using electrical signals obtained from chemical reactions so it's possible observe and identify scalp's electrical potentials variations and retrieving information from the associated brain waves; this is a safely process because EEG is a passive measuring device and less invasive than other methods.

This paper describes a set of guidelines for the affective level design and manipulation of emotions in a Role Playing Game. We investigate how these can be applied to practical design rules and then make use of a BCI to experimentally test the users response. In this way we try to extract objective data from subjective emotions, and use these to uniquely connote the tasks and the areas developed for the different game levels.

In Section II recent related work, which use videogames to study player's behaviors are reported. Section III proposes the formal guidelines and their practical application into game level design, while Section IV briefly illustrates the technical tools used. Section V describes the experimental setup, while Section VI shows and comments the results coming from the data analysis. Finally conclusions and suggestions for future work are drawn in Section VII.

II. Related work

Nacke *et al.* [6], [4] were the first to formalize game level design guidelines and to employ BCI to evaluate them, using the first-person shooter (FPS) *Half-Life*; also in Nacke and Lindley [7] the opportunity to deal with the events and tasks of a game level is considered.

Plotnikov *et al.* [8] investigate the brain waves for the concepts of boredom and flow using the *Tetris* videogame; in the work of Nijholt *et al.* [9] there is an extensive survey on current BCI research about HCI studies with the viewpoint of games and game design, featuring suggestions and proposals.

Herrewijn *et al.* [10] use surveys to analyze the factor of "Immersion" in the videogame *Fallout:New Vegas* while Reuderink *et al.* [11] voluntarily cause frustration to the gamers with an adapted clone of *Pacman* also analyzing their EEG data; in Canossa *et al.* [12] there is a dissertation about metrics to discover the frustration factors during a gaming session using internal game data. The EEG data are also

used in Lee and Tan [13] to classify the tasks depending on their cognitive and mnemonic effort with the videogame *Halo*, while Smith *et al.* [14] focus the analysis on the interfaces, to improve their usability with complex tasks. Salminen and Ravaja [15] use the videogame *007:NightFire* to inquire about the virtual violence.

A BCI headset can be used not only to gather neurological data, but also as a real-time input device: in Pour *et al.* [16] the headset is used to control the interaction in a clone of the videogame *Breakout*.

III. LEVEL DESIGN FOR AFFECTIVE LUDOLOGY

The design of a game and its levels depends on the *gameplay*, i.e. the set of "internal rules" that characterize the specific game dynamics and so denote particularly every game genre (FPS, platform, puzzle,...). According to the definition provided by Brown and Cairns [17] there are three phases of participation (engagement, engrossment, total immersion) in which the concept of immersion and flow are very close. Nacke *et al.* [6] provided a set of affective design guidelines and developed three game levels with the distinction between Immersion and Flow by dividing the concept of environment from that of combat which constitutes the focus of the FPS gameplay.

When considering RPG gameplay features, the focus moves from just action and shooting challenges to other different aspects:

- the interaction using dialogues
- the presence of allies in a group called "party"
- the developing of moral choices and riddles

The gameplay of an RPG is less immediate than FPS because it's based on statistics, bird-flight view, objects inventory and character's skills development, with a great importance given to the environment's exploration to reach the "sense of wonder". For these reasons immersion can be treated as a necessary condition to reach the flow emotional state ($Immersion \subset Flow$). In order to stress this evolution of the participation, only two game levels have to be developed for Neverwinter Nights 2: one will focus on boredom, while the other will stress flow.

A. Boredom game level

Fisher [18] defines boredom like a transitory unpleasant affective state with lack of concentration and difficulty during a task, Csíkszentmihályi [19] furthermore denotes it like a state in which player's skills are greater than required. Formal design guidelines:

Guideline 1: Assuming a spatial progression point p to indicate game level progression and given a set E of n types of enemy e_a with constant strength a smaller than player strength $b > a$, then the player challenge function $f_{ch}(p)$ for an encounter k with enemies type e_a at progression point p should remain constant $\forall e_n \in E$.

Guideline 2: Given n textures $t \in T = \{t_1, t_2, ..., t_n\}$ in a game level L, then L becomes less interesting if contains less elements of T so that $L = \{t_k, ..., t_m\} \subset T$, $m < n$;

Fig. 1. The tasks "text dialogue" and "single fight" in the boredom game level.

the same should be for all other structural and visual features like sounds, animations, light and visual effects, trees, enemy types, items, 3D models, weapons.

Guideline 3: For each progression unit p the player should be constantly rewarded with $n = p$ items with i_h that refreshes player health value and i_a for the ammunition value so that the full player ammunition supply value $A = \sum_{a=0}^{n} i_a$ and the full player health value $H = \sum_{h=0}^{n} i_h$ are often reached.

Guideline 4: No real winning condition C_k so that $C = \emptyset$ and the victory goal $V = \{C_1, C_2, ..., C_n\} = \emptyset$ doesn't give any reward r_v to the player.

Guideline 5: Given n characters a with a set S of skills belonging to a specific class t which can be in $C =$ {warrior, wizard, thief, animal, unskilled}, if a_t can join the player's party G becoming his ally $a_t \in G = \{a_{t1}, a_{t2}, ..., a_{tn}\}$ in a game level L, then L becomes less interesting if contains much less elements of G so that $L = \{a_{tk}, ..., a_{tm}\} \subset G$, $tm \ll tn \vee \exists a_t \ t = unskilled$.

Guideline 6: Given n generic dialogues $d \in D = \{d_1, d_2, ..., d_n\}$ that a player can have with the NPCs of the game level L, then L becomes less interesting if contains much less elements of D so that $L = \{d_k, ..., d_m\} \subset D$ $m \ll n$; given also n questions $q \in Q = \{q_1, q_2, ..., q_n\}$ that the player can choose during an articulated dialog $d_a \in D$, assuming that the relative answer to q_k is a_k, then L becomes less interesting if the dialogues contain much less elements of Q so that for the most part of D is $d_a = \{q_k, ..., q_m\} \subset Q$, $m \ll n$.

Referring to the previous formal guidelines, three areas have been designed by editing the *2311_tunnels* level, contained in the *2300_Crossroad_Keep_Adv* package of the original game, because it has an environment with linear paths, the prevalence of gray and blue colors, few structures and NPCs (Fig. 1).

The challenge offered by this game level is minimal with constant difficulty: there are only two types of weapons and enemies (with low level skills of resistance and attack) and player's health and weapon's munitions are always at the maximum, since every enemy releases an health potion at his death and every chest releases munitions and weapons; the story plot is really basic with a very simple task to accomplish, without victory conditions or gratifications.

The structural and visual features are poor with only indoor environments to emphasize the linearity and repetitiveness of the areas (the path in the third area is the reverse of the one in the first area); there are no visual effects or music and

sounds for battle and environment, and every NPC has the same 3D model. The allies are fundamentally unnecessary and tedious, consisting in an unskilled not interactive human character, without Artificial Intelligence, which only follows the player. Dialogues are performed only in "text version" without animations or camera changes: they are unbranched and useless, filled with out-of-context sentences and little variety of questions and answers that don't add anything to the story plot; furthermore some of them create false expectations in the player.

B. Flow game level

In the field of (video)games the notion of flow is closely related to that of *fun* even if doesn't coincide with it: Nakamura and Csíkszentmihályi [20] observed the behaviors of chess players and sport players noting that their enjoyment derived by the mere fact to accomplish their activity independently by other types of rewards; the flow is intended like an emotional peak with total involvement and mental attention in which the skills fully meet the challenge of a task and the player is fully absorbed by it. Formal design guidelines:

Guideline 1: Given a set of indoor level parts $I \neq \emptyset$ and a set of outdoor level parts $O \neq \emptyset$, the game level L should be a set union of outdoor and indoor level parts $L = \{I, O\}$.

Guideline 2: Assuming a spatial progression point p to indicate a game level progression and given a set E of n type of enemy e_a with strength a, then the player challenge function $f_{ch}(p)$ for an encounter k with an enemy e_a at progression point p has to progressively increase $\forall e_a \in E$ (Ermi and Mäyrä [21]).

Guideline 3: Given n sets $X_k \neq \emptyset$ of different structures in a game level L like textures X_t, lighting effects X_l and sounds X_s, then L becomes more atmospheric and fosters imagination if $L = \{X_t, X_l, X_s..., X_n\}$; the same should be for all other structural and visual features like animations, fire effects, trees, enemy types, items, 3D models, weapons.

Guideline 4: For a player progression unit p, after a set progression interval p_i for n intervals in a game level $L = \{p_1, p_2, ..., p_n\}$, a reward type r_p from a set of n rewards $R = \{r_1, r_2, ..., r_n\}$ should be given to the player such as ammunition, health packs, money, new weapons and spells.

Guideline 5: There's at least one winning condition C_k of the victory goal $V = \{C_1, C_2, ..., C_n\}$, $V \neq \emptyset$; the achievement of V must lead to one or more rewards r_v from a set of n rewards $R = \{r_1, r_2, ..., r_n\}$ in order to gratify the player's efforts.

Guideline 6: Given n characters a with a set S of skills belonging to a specific class $t \in C =$ {warrior, wizard, thief, animal, unskilled}, if a_t can join the player's party G becoming his ally $a_t \in G = \{a_{t1}, a_{t2}, ..., a_{tn}\}$ in a game level L, then L becomes more interesting if $L = \{a_{tk}, ..., a_{tm}\} \subseteq G$ with $G \neq \emptyset \land \forall a_t \in G$ $t \neq unskilled$.

Guideline 7: Given n generic dialogues $d \in D = \{d_1, d_2, ..., d_n\}$ that a player can have with the NPCs of the game level L, has to be $L = \{d_k, ..., d_m\} \subseteq D$; given also n questions $q \in Q = \{q_1, q_2, ..., q_n\}$ that the player can choose

Fig. 2. The tasks "animated dialogue" and "group fight" in the flow game level.

during an articulated dialog $d_a \in D$, assuming that the relative answer to q_k is a_k then the game level L becomes more interesting if for the most part of D is $d_a = \{q_k, ..., q_m\} \subseteq Q$ and d_a is a *narrative or ludic game level structure* useful for reaching a condition of victory V or to achieve useful information and items.

Based on these guidelines, four areas have been designed by editing the *3010_highcliff*, *3000_castle_never*, *3063_merchant* and *3032_th_canyon* levels contained in the *3000_Neverwinter_A3* package of the original game. Linking the areas, the original enemies, paths, quests and dialogues have been replaced to create a new coherent story plot; three original dubbed dialogues are preserved (Fig. 2).

The game level must be characterized by the balance between difficulty and player's skills, in fact if the challenge is higher than the ability the activity becomes overwhelming and generates anxiety, while if the challenge is lower than the player's ability, it provokes boredom. In order to continue to experience flow, a person must continue to learn new skills [22]: the challenges steadily increase due to the variety of enemy types provided by every area (better statistics, weapons, and skills). The enemies can also attack in groups with ambushes, furthermore the previous encountered enemies grow and become harder to defeat.

Environment structures are diversified with indoor and outdoor areas, each of which has a specific setting (cliff, castle, haunted town, volcanic caves) and some of them offer optional paths. The structural and visual features are rich of 3D models, items, effects (lights, animated water and trees), particular music and sounds with alternation between day and night and weather conditions.

Rewards are balanced with the player's skills and the challenge offered, in fact only few enemies (the harder to defeat) release money or health potions at their death, and the chests which release magic objects, ammo and weapons are located only before or after the more challenging combat sections; there are wearable objects which increase player's skills which can also grow using the experience points that the character gains. Within the different types of weapons and equipment, also spells are included.

The story plot is rich, with moral and emotional entanglements and well-defined and satisfactory goals (a rescue,

an enrollment, a discovery) and related rewards; some tasks are also linked to the player's choices and influence the prosecution of the story. There is a party of skilled allies, useful and diversified by their particular skills (a wizard, a warrior, a thief) with different ways to enlist them (automatic, payment, dialogue). All the allies in the party are interactive with a proactive Artificial Intelligence also additional non-interactive NPCs (guards and peasants) may help the player to perform some tasks.

Dialogues are useful and articulated: they use animations and shots and provide useful information for the plot (memories of events, opinions, riddles, narrative elements). Moreover, the conclusion of the quest (positive or negative) depends on the moral choices (greed, compassion) made in some of them; some dialogues have professional voice dubbing which increases player's emotional involvement.

IV. The technologies

The Emotiv EPOC Headset is a cheap neuro-signal acquisition and processing wireless headset with 14 wet sensors and 2 reference sensors which is capable of detecting brain signals: the Control Panel provides an interface which allows monitoring in real-time the headset's signals implementing the *Affectiv Suite* for visualizing in charts the five detected emotions (Engagement, Frustration, Meditation, Excitement and LongTermExcitement).

The choice of using Neverwinter Nights 2 for the game level design was made considering several factors:

- the *RPG* gameplay which offers variety of story plot, characters and events
- a complete editor (*Electron*) used even by the authors which provides visual tools to create dialogues, NPCs (Non-Player Characters) with Artificial Intelligence, environments, triggers, objects and visual effects
- the powerful C-like scripting language *NWScript* which allows to manage internal game's data and variables
- the possibility to customize the game interface with XML

For the experimental phase there are two types of data to extract during the gaming session:

- internal: environmental values, character's values, map coordinates
- external: player's affective data which come from the headset

The data sources must be automatically stored and synchronized with a timestamp to facilitate their analysis. The open source C++ tool NWNX[23] allows the implementation of a *plugin* which manages the interchange of data between the game, the screen and the "outside world": developed for the online gaming, it uses the server version of the game and permits the calling of the plugin's code from a NWScript and its integration with the code libraries of the headset's SDK .

Screen captures of the gaming sessions can be also useful for the experiments so an *XML UIPane* widget will enable NWNX printing in real-time on the game interface the same timestamp used to synchronize the internal data.

V. Experimental study

Laboratory experimentation involved 13 volunteers male academic students tested in a range hour between 10.00 AM and 07.00 PM on weekdays. A single session was about one hour and half in which the subject had to play the two game levels wearing the headset, and every gaming session was recorded using a screen capture software.

At the beginning of a session a brief explanation about the modalities of the study was provided, followed by the compilation of an informed consent and of a pre-questionnaire about subject's gaming preferences; every subject played using the same game character (a neutral half-elf warrior with all the character statistics set to an average value).

Next a step-by-step game tutorial consisting of two areas with brief tasks was presented, with the aim of learning commands and game rules familiarizing with the user interface. Some of the offered tasks were: "move the character", "rotate and zoom the screen view", "have a chat with someone", "take the control of an ally", "open the inventory or a chest and take an item", "launch a spell", "fight", "rest to recover health".

After the tutorial, the headset was put on subject's scalp with a phase of signal tuning and the starting of the data gathering software. Following the first gaming session the subject compiled a post-questionnaire about his gaming experience. The next game level was then played with the compilation of a new post-questionnaire and finally the headset was removed from the subject.

Different data sources were available for every gaming session:

- an MPEG file with the gaming session screen capture
- the pre and post questionnaires
- two log files with data coming from the game and from the headset, synchronized by a timestamp

Every entry of the log for the *events data* (Fig.3-1) contains the timestamp and the tag of the meaningful event occurred during the gaming session captured using internal NWNScript game triggers.

Every entry of the log for the *affective data* (Fig.3-2) contains the timestamp, the name of the area in which the player was with its coordinates (X and Y values) and the punctual values of the emotions that the headset recognized: these ones are in a range between 0 and 1 and classified by the internal algorithms of the Emotiv EPOC.

A schematization of the emotions available to understand the affective data is the Circumplex model of affect (Fig. 4) by Russell *et al.* [24], [25]: in this two-dimensional spatial model every human affect could be defined on a circumference combining the bipolar dimensions of *valence* and *arousal*; the first suggests if an affect is pleasant or unpleasant (X axis), the second suggests the activation or the deactivation of an emotion (Y axis).

The *Engagement* denotes an increment of the beta brain-wave and a reduction of the alfa ones [26]: by the Emotiv EPOC headset it's associated to the mood states of participation and attention and as an indicator of fun so if the aim
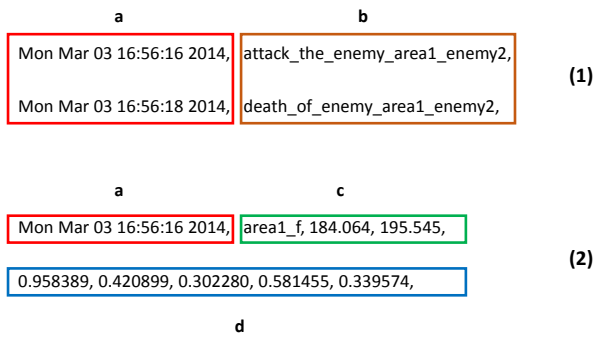
Fig. 3. (1): part of the events log with a sequence of entries featuring timestamps (box "a") and the related events occurred during the gaming session (box "b"). (2): an entry line (divided in two) of the affective log featuring the timestamp (box "a"), the character's map coordinates (box "c") and the punctual values for the five recognized emotions (box "d").
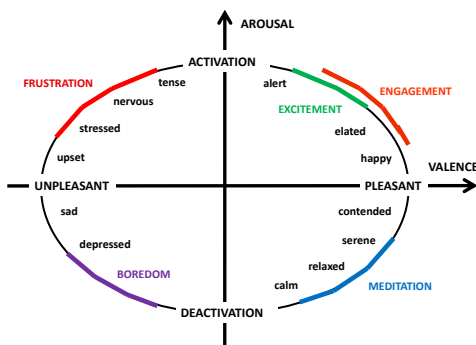

Fig. 4. An interpretation of the circumplex model of affect: Excitement is pleasant and activated so it is located in the upper-right quadrant like Engagement which is its extension; Meditation is pleasant but not activated, so it is located in the bottom-right quadrant. Frustration and Boredom are unpleasant, but the first is activated so it is located in the upper-left quadrant, while the second is in the bottom-left.

was to induce the flow state, most of the tasks and events must show an increase of it for the player's interest perceived; because *Boredom* is a deactivated emotion it can be seen as the opposite of engagement so if a game level was designed to induce it, most of the tasks and events should show a decrease of engagement.

The *Excitement* is a positive indicator linked to a short-time emotional peak so it may be unfit as an indicator for flow or boredom: it is better to consider the *LongTermExcitement* like a global value which expresses how stable over time the excitement is. As observed for engagement, if the aim was to induce boredom most of the tasks and events should show a decrease while the opposite should be true for the flow state. Engagement and Excitement are closely related to the concept of flow due to the mutual interaction between interest and involvement while facing the game tasks.

Inducing and analyzing *Frustration* and *Meditation* in game tasks is difficult because they have a negative valence and are influenced by subjectivity, for example different players can perceive the same amount of frustration as a challenge or as

	a1	a2	a3	Tot.
subj.1	4	5	3	12
subj.2	5	6	4	15
subj.3	5	6	4	15
subj.4	5	6	4	15
subj.5	5	6	4	15
subj.6	5	6	4	15
subj.7	5	6	4	15
subj.8	5	7	4	16
subj.9	5	5	3	13
subj.10	5	6	4	15
subj.11	5	6	3	14
subj.12	5	8	4	17
subj.13	5	6	4	15
Tot.	64	79	49	192

a reason to quit the game.

VI. EVALUATION OF THE EEG DATA

A game level can be considered as a sequence of challenges and situations and more specifically as the sequence of tasks and events which the player faces during his gaming session; for the evaluation it was decided to consider the tasks and the events of the areas defined in the design phase: this approach allows the interpretation of data that greatly changes over time due to the duration of a gaming session, focusing on factors like repetitiveness, plot development and modifications in the environment.

The significant tasks and events counted for the subjects are those defined in the design phase extracted from the log files: *dubbed dialogue*, *riddle dialogue*, *dialogue*, *attack an enemy or a group of them* (fight), *death of an enemy or his entire group* (fight end), *chest opened* (with its items gathered); there are two further events which are annotated using the recoded video: *character growth* (skills upgrade) and the *stealing task* with the thief character.

Because the affective data was recorded 10 times per second it is important to identify and extract the sequential data entries between the beginning and the end of a task using the timestamps, with the aid of the video recording and the events log to identify the time edges.

Considering a player session, the subject may face very different task and events (TableI and II): the boredom game level presents an average of 14.7 events while the flow level offered about the double of them (27.9) due to the different structure and development guidelines; some type of event are not present in the boredom level, like dubbed and riddle dialogues or the skills upgrade and the stealing action.

As seen in the design Section, tasks and events can be very different even if they are of the same type, in fact they depend

TABLE II
SUBJECT'S EVENTS COUNTS BY AREA: FLOW GAME LEVEL

	a1	a2	a3	a4	a5	Tot.
subj.1	6	9	13	4	0	32
subj.2	6	8	6	5	0	25
subj.3	6	10	13	5	0	34
subj.4	7	9	7	4	1	28
subj.5	6	10	8	3	1	28
subj.6	7	8	6	5	0	26
subj.7	7	9	7	1	0	24
subj.8	7	9	12	0	0	28
subj.9	7	9	7	5	0	28
subj.10	6	10	14	4	0	34
subj.11	6	9	6	4	1	26
subj.12	5	10	12	5	0	32
subj.13	6	8	7	0	0	21
Tot.	82	118	118	45	3	366

TABLE III
BINARY CLASSIFICATION RESULTS FOR THE TWO GAME LEVELS

	boredom	flow
original events	366 (ext.)	366
classified events	423	309
correctly classified	298	241
accuracy	81.42%	65.85%
miss rate	18.58%	34.15%
false positive rate	29.55%	22.00%
generalization error	0.2896	
bias	0.1656	

on the kind of emotion which they want to manipulate and finally by the game level (and area) in which they are located.

In order to evaluate the game level design we chose to perform a supervised machine learning binary classification with the clauses of *boredom* and *flow*: in this way it is possible to investigate if the brain-recorded data is characteristic enough to differentiate among the two set of development guidelines.

The data for training is a set of points (vectors) x_i along with their categories y_i; for some dimension d, the $x_i \in R^d$ and the $y_i = \pm 1$. The equation of a hyperplane is

$$< w, x > + b = 0 \qquad (1)$$

where $w \in R^d, < w, x >$ is the inner (dot) product of w and x, and b is real.

The following problem defines the best separating hyperplane: find w and b that minimize $||w||$ such that for all data points (x_i, y_i)

$$y_i(< w, x_i > + b) \geq 1 \qquad (2)$$

The support vectors are the x_i on the boundary, those for which $y_i(< w, x_i > + b) = 1$.

The primary data features associated to a game level task are the emotive data retrieved by the headset, but considering the entire duration of a task measures are required which summarize the values in the log: for every affective emotion E, each of which consists of a set of n log entries related to the task, we considered:

- the values of the *angular coefficient* β and the *intercept* α from a linear regression line calculated considering *time* (T) as a positive variable which constantly increases

$$E_i = \alpha + \beta T_i + \mu_i, i : 1..n \qquad (3)$$

where μ_i is the statistical error

- the *Pearson product-moment correlation coefficient* considering *time* (T) in order to estimate its relationship with an *affective emotion* (E)

$$\rho_{T,E} = \frac{cov(T, E)}{\sigma_T \sigma_E}, -1 \leq \rho_{T,E} \leq +1. \qquad (4)$$

In particular the *Pearson correlation coefficient* r_e calculated for a sample of n affective values in the task's time interval considered:

$$r_e = \frac{\sum_{i=1}^{n}(T_i - \bar{T})(E_i - \bar{E})}{\sqrt{\sum_{i=1}^{n}(T_i - \bar{T})^2}\sqrt{\sum_{i=1}^{n}(E_i - \bar{E})^2}},$$

$$-1 \leq r_e \leq +1 \qquad (5)$$

if the outcome of Eq. 5 is a strongly significant value ($r_e > 0.7$ or $r_e < -0.7$), then we can assume evidence of a local correlation for the selected emotion; a positive coefficient detects a direct correlation while a negative detects an inverse correlation).

- the arithmetic *mean* $\bar{E}$
- the *variance* σ_E^2

Considering the 13 subjects we have a 558x25 matrix M in which there are 588 labeled events (192 b of the boredom level and 366 f of the flow) with 25 numeric features (five for every emotion); since the set is unbalanced, the number of boredom cases was increased by randomly sampling 174 cases from the original cases.

The feature data matrix becomes 732x25 and it is used as the matrix of predictors to train a SVM (support vector machine) classifier: while by default the training uses a 10-fold cross validation a further training has been done using the option "leaveout" which trains the system n times iteratively using $n - 1$ cases and testing it with the one left out; after the new training session the system has been tested with the same matrix used to train it with the results in the Table III. The evaluation of all this measures by the classifier will suggest if a task was well-designed to induce some trend variations in the manipulation of the emotion also considering the natural presence of noise in the data.

The *accuracy* of the classification is in both game levels up to 65% suggesting a satisfactory success for the design phase: the presence of an hyperplane which splits the events

TABLE IV
MOOD-CONGRUENT EVENTS BY AREA: BOREDOM GAME LEVEL

	total number (ext.)	correctly classified	rate
area1	125	115	92.00%
area2	155	124	80.00%
area3	86	59	68.60%
Tot.	366	298	

TABLE V
MOOD-CONGRUENT EVENTS BY AREA: FLOW GAME LEVEL

	total number	correctly classified	rate
area1	82	26	31.71%
area2	118	75	63.56%
area3	118	97	82.20%
area4	45	40	88.89%
area5	3	3	100%
Tot.	366	241	

TABLE VI
MOOD-CONGRUENT EVENTS BY TYPE: BOREDOM GAME LEVEL

	total number (ext.)	correctly classified	rate
dialogue	158	135	85.44%
single fight	111	82	73.87%
chest opened	97	81	83.50%
Tot.	366	298	

in the geometric space of the feature highlights a strong characterization between the measures of the tasks of the same type with the 65.85% for the flow game level tasks and the 81% for the boredom (even considering the extension of its training cases). The false positive classification is under 30% for both the game levels with a greater value for the boredom (29.55%).

A. Classification for game level areas

Another useful aspect to evaluate the design of the game level is to consider to which game level area the correctly classified tasks belong, comparing their number with the original; the results for the areas of the boredom game level are in Table IV while for the flow in Table V. For the three areas of the boredom game level the accuracy is up to 65% with a peak for the area 1 (92%) which suggest that this area and the area 2 (80%) were well-designed by the guidelines because they have events and tasks highly characterized by their recognized emotive values: while the first two areas are the same with repetitiveness of path and visual structures the area 3 introduces some small variations which however are at least sufficient to reduce the boredom perceived by the subjects.

For the three areas of the flow game level the accuracy is variable: over the 80% for area 4 (88.89%) and area 3 and smaller for area 2 (63.56%) and area 1 (31.71%): the progression of the values suggests the growth of the subject's engagement and long-term excitement due to the development of the plot story and the challenges faced with increasing new allies and skills.

The fourth area is the shortest but results to be the best-designed due to the presence of duels, dubbed dialogues, a moral choice and a great enemy like a dragon while the third area (which is the largest) has a moral choice and an optional path in which recruiting the thief character (to perform the stealing task). The fifth area was the final stage containing only one dialogue with a 100% rating but since was reached only by 3 subjects it is not very significant for the evaluation.

B. Classification for task and event types

Using the log files it is possible to label every task faced by a subject: it can be useful to inquire about the types of task and events and if there are types classified better than others; the results for the boredom game level are in Table VI while for the flow are in Table VII. In the boredom game level, *dialogues* are the best recognized type (85.44%) followed

by the *chest opened* with its items gathered by the players (83.50%). The classification rate is over 70%, evidence of a correct development phase and sound guidelines to increase the poorness of dialogues (performed only in texual mode) and the repetitiveness of the chest's items founding.

Also in the flow game level the dialogues are the most recognized: the *dubbed* (93.94%) and the *riddle* (84.61%) have been developed specifically for the gameplay of the flow (as the *skills upgrade* phase also up to 70%) and so their classification success justifies the goodness of the guidelines proposed in the design phase; the most frequent types are *dialogue* (articulated and performed using animations and shoots) and *single fight*, with a rate satisfactorily up to 64% as *fight vs. a group* which are less numerous.

The difference between the two type of fights (single duel or against a group of enemies) is about 0.50% so there is no justification for their differentiation in the guidelines; the *chest opened* task is the only one under 50% of classification and,

TABLE VII
MOOD-CONGRUENT EVENTS BY TYPE: FLOW GAME LEVEL

	total number	correctly classified	rate
dubbed dialogue	33	31	93.94%
riddle dialogue	13	11	84.61%
dialogue	110	72	65.45%
single fight	107	69	64.48%
fight vs. a group	25	16	64%
chest opened	63	30	47.62%
skills upgrade	11	8	72.73%
stealing action	4	4	100%
Tot.	366	241	

comparing it with the value for the boredom level (-35.88%), it becomes evident that the repetitiveness is more efficient than the variety of the retrieved items. The *stealing action* task was optional and performed only by 4 subjects, but it was characterized very well resulting in a rating of 100%.

VII. CONCLUSIONS

The results from experimentation are very encouraging about the usefulness of the proposed guidelines and underlines the goodness of the design phase of the two game level both for their structure (the areas) and for the implementation and conception of the tasks and events; the dialogues, the fulcrum of the RPG gameplay, prove the ability to manipulate the player's emotions positively (in the flow game level) and negatively (in the boredom game level). The factor of repetitiveness and progression (of skills and challenges), respectively at the base of the guidelines of boredom and flow, become evident considering the results about the areas classification.

In future works the analysis of the raw data provided by the headset's sensor will be very useful, in order to extract the trends of the single brain waves and compare them with the affective classification detected from the internal algorithms used there; also the different questionnaires require analysis.

The introduction of some forms of DDA (dynamic difficulty adjustment) could benefit from the results of this work: with a real-time classification of the emotions it could be possible to change the trend of an emotion during a gaming session.

The screen capture of the gaming session can be (automatically) analyzed to detect the dynamics of gameplay like strategies and behavior patterns of the gamers; also the AI of the allies can became dynamic and benefit from a mechanism of machine learning based on this patterns. Finally, given the extreme variety of videogame genres, it can be useful to test the guidelines with other games presenting a different gameplay (survival, stealth, ...) or different modalities of affective interaction (on-line with real people as ally and enemies) like MMORPGs.

REFERENCES

[1] J. P. Gee, "Learning and games," *The ecology of games: Connecting youth, games, and learning*, vol. 3, pp. 21–40, 2008.

[2] G. Frasca, "Ludologists love stories, too: notes from a debate that never took place." in *DIGRA Conf.*, 2003.

[3] M. Csikszentmihalyi, "Does being human matter? on some interpretive problems of comparative ludology," *Behavioral and Brain Sciences*, vol. 5, pp. 160–160, 3 1982. [Online]. Available: http://journals.cambridge.org/article_S0140525X00010980

[4] L. Nacke, "Affective ludology: Scientific measurement of user experience in interactive entertainment," Ph.D. dissertation, Blekinge Institute of Technology, Karlskrona, Sweden, 2009, http://phd.acagamic.com, (ISBN) 978-91-7295-169-3.

[5] R. L. Mandryk and K. M. Inkpen, "Physiological indicators for the evaluation of co-located collaborative play," in *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. ACM, 2004, pp. 102–111.

[6] L. Nacke, S. Stellmach, and C. Lindley, "Electroencephalographic assessment of player experience: A pilot study in affective ludology," *Simulation and Gaming*, vol. 42, no. 5, pp. 632–655, 2011, cited By 5. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-81755163956&partnerID=40&md5=d5b1a5426fb88f962d4a2feacb322ed3

[7] L. Nacke and C. A. Lindley, "Flow and immersion in first-person shooters: Measuring the player's gameplay experience," in *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, ser. Future Play '08. New York, NY, USA: ACM, 2008, pp. 81–88. [Online]. Available: http://doi.acm.org/10.1145/1496984.1496998

[8] A. Plotnikov, N. Stakheika, C. Schatten, F. Belotti, D. Pranantha, R. Berta, and A. De Gloria, "Measuring enjoyment in games through electroencephalogram (eeg) signal analysis," in *Proceedings of the 6th European Conference on Games-Based Learning (ECGBL 2012)*, 2012, pp. 393–400.

[9] A. Nijholt, B. Reuderink, and D. Oude Bos, "Turning shortcomings into challenges: Brain-computer interfaces for games," in *Intelligent Technologies for Interactive Entertainment*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, A. Nijholt, D. Reidsma, and H. Hondorp, Eds. Springer Berlin Heidelberg, 2009, vol. 9, pp. 153–168. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-02315-6_15

[10] L. Herrewijn, K. Poels, and G. Calleja, "The relationship between player involvement and immersion: An experimental investigation." in *FDG*, 2013, pp. 364–367.

[11] B. Reuderink, A. Nijholt, and M. Poel, "Affective pacman: A frustrating game for brain-computer interface experiments," in *Intelligent Technologies for Interactive Entertainment*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, A. Nijholt, D. Reidsma, and H. Hondorp, Eds. Springer Berlin Heidelberg, 2009, vol. 9, pp. 221–227. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-02315-6_23

[12] A. Canossa, A. Drachen, and J. R. M. Sørensen, "Arrrgghh!!!: Blending quantitative and qualitative methods to detect player frustration," in *Proceedings of the 6th International Conference on Foundations of Digital Games*, ser. FDG '11. New York, NY, USA: ACM, 2011, pp. 61–68. [Online]. Available: http://doi.acm.org/10.1145/2159365.2159374

[13] J. C. Lee and D. S. Tan, "Using a low-cost electroencephalograph for task classification in hci research," in *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '06. New York, NY, USA: ACM, 2006, pp. 81–90. [Online]. Available: http://doi.acm.org/10.1145/1166253.1166268

[14] M. E. Smith, A. Gevins, H. Brown, A. Karnik, and R. Du, "Monitoring task loading with multivariate eeg measures during complex forms of human-computer interaction," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 43, no. 3, pp. 366–380, 2001.

[15] M. Salminen and N. Ravaja, "Increased oscillatory theta activation evoked by violent digital game events," *Neuroscience letters*, vol. 435, no. 1, pp. 69–72, 2008.

[16] P. A. Pour, T. Gulrez, O. AlZoubi, G. Gargiulo, and R. A. Calvo, "Brain-computer interface: Next generation thought controlled distributed video game development platform," in *Computational Intelligence and Games, 2008. CIG'08. IEEE Symposium On*. IEEE, 2008, pp. 251–257.

[17] E. Brown and P. Cairns, "A grounded investigation of game immersion," in *CHI'04 extended abstracts on Human factors in computing systems*. ACM, 2004, pp. 1297–1300.

[18] C. D. Fisher, "Boredom at work: A neglected concept," *Human Relations*, vol. 46, no. 3, pp. 395–417, 1993.

[19] M. Csikszentmihalyi, *Beyond boredom and anxiety*. Jossey-Bass, 2000.

[20] J. Nakamura and M. Csikszentmihalyi, "The concept of flow," *Handbook of positive psychology*, pp. 89–105, 2002.

[21] L. Ermi and F. Mäyrä, "Fundamental components of the gameplay experience: Analysing immersion," *Worlds in play: International perspectives on digital games research*, vol. 37, 2005.

[22] D. O. Broin, "Using a criteria-based user model for facilitating flow in serious games," in *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2011 Third International Conference on*. IEEE, 2011, pp. 63–69.

[23] *Neverwinter Nights eXtender v.4*, http://www.nwnx.org/.

[24] J. A. Russell, "A circumplex model of affect," *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[25] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and psychopathology*, vol. 17, no. 03, pp. 715–734, 2005.

[26] *Emotiv EPOC Software Development Kit and User Manual for Release 1.0.0.5, par. 3.4.3*, wiki.emotiv.com/tiki-index.php?page=Documentation.