FACIAL ANIMATION AND MOTION CAPTURE: KEY ROLE FOR THE COMMUNICATION

Sergi VILLAGRASA FALIP
Grup de Recerca en Tecnologies Mèdia, Enginyeria i Arquitectura La Salle, Universitat Ramon Llull, Barcelona, 08022, Spain

Jaume DURAN CASTELLS
Didáctica de la Educación Visual y Plástica - Comunicación Audiovisual, Universitat de Barcelona, Barcelona, 08035, Spain

David FONSECA ESCUDERO
Grup de Recerca en Tecnologies Mèdia, Enginyeria i Arquitectura La Salle, Universitat Ramon Llull, Barcelona, 08022, Spain

ABSTRACT

In the last 100 years we have observed a great development of the technology. This development applied to the mass media as the radio, the cinema, the television, or Internet have influence in the human communication and the interaction of the society with the technologies.

The new communicative environments need new actors (avatars) with high communicate face expression.

The modeling and the animation of an expressive human face are the greatest challenges in computer graphics because we are able to detect the nuances of even the most basic movement of the face.

The cinema and the video games are the fields that more have worked the creation of virtual, "real" and believable characters. With this work, we have analyzed the concrete case of the 3D animation cinema, studying how the technologies of character's creation (with or without motion capture technique) can generate a major empathy with the user.

Nowadays, the animation technique called motion capture is usually present in the production of animation cinema. Our main objectives are to evaluate if these skills can improve the facial animation and if this progress generates amplified emotional response on users.

Keywords: 3D facial animation, motion capture, communication technology, empathy.

1. INTRODUCTION

When a user says "I love or enjoy this movie" or "I like this character" it can mean many different things. The kind of movie or the technical characteristics of the character that are being evaluated can change the experience of enjoying a film. The user constructs a parallel story when he is watching a film based on his knowledge and social environment (acquired from his real experience, e.g. schemas and stereotypes) and his experiences with other fictional and non-fictional narratives [6, 10].

The audience has emotional reactions to situations and events that occur in the films [9, 23, 24], and the emotional reactions to entertainment content have been the focus of much theory and research [3, 16, 17, 18].

Perceived realism of mass media content is often defined as "the degree of similarity between virtual and real-life characters and their behaviours" [23]. We need to research if the realism in the facial animation is a main item to obtain a successful connection between the society and the cyber-technologies.

The studies about the capture of the movement and the animation began throughout 19th Century and were developed throughout 20th Century. However, the relation between these two concepts will not have a great meaning until a few years ago, with the arrival of the motion capture in the computer animation. In this study, we will describe a brief perspective and its evolution, since Edward Muybridge o Étienne Jules Marey until actuality and we will study the empathy of characters that are being developed with motion capture in front of animation characters without this technology in relation with the user.

With the motion capture technology (mocap), the work of animation can develop characters in a reduced time; it contributes for better interpretation of the characters, and improves the expressiveness of the bodies. This technology contributes to generate 3D characters with high reality in front of the "classical" animation characters, also known as "cartoons".

Nowadays, we can divide the capture in two types: the body and the face. Facial animation is still an immature field, with undefined methodology. The human face is an extremely complex system that is very difficult to model and the researchers are looking for methods to describe and to model the human movements.

The aim of the research reported in this paper is to explore the evolution of the animation cinema and the implementation of mocap to generate 3D characters. We will evaluate six scenes from six animation films to identify the dimensions of movie enjoyment and empathy of the user with the main character studied.

Our main hypothesis is that a character developed with mocap has more empathy to the user, because the user identifies with the expressions and the facial movements. This view of enjoyment is closely associated with mood-management theory [25].

To develop our study, we will focus to evaluating three main levels:

- Sympathy related to the character.
- Narrative Immersion: estimation of the level of immersion and identification of the user with the character.
- Narrative Realism: perception incoherence's in the evaluation of the quality and character behaviour.

2. THE IMAGE IN MOVEMENT AND THE ANIMATION

In 1874, Pierre-Jules Janssen (1824-1907) registered for the first time a photographic sequence of a natural phenomenon with one device: Venus's traffic opposite to the Sun from Yokohama (Japan) with 48 images in 72 seconds. Two years before, Eadweard Muybridge (1830-1904) was photographing the trot of a horse to demonstrate if in the course of the gallop, this one was touching or not with four legs in the ground. But he had to wait until 1878. In summer of this year, finally, after diverse attempts, in Palo Alto (California), Muybridge realized the first photographic sequence of an animal in movement.

The experience allowed other inventors to accelerate the search of a system to realize and to project photographic images in movement, and to study the movement of the animals and people. Étienne Jules Marey (1830-1904), in 1882, invented the photographic gun, to obtain with an exhibition of 1/720s, twelve images of the flight of a bird, which it allowed him later, and seeing that the above mentioned images were small and little confirmed, to develop the first "chronophotographies" with a fixed plate, where thanks to a camera with a shutter that was circulating to great speed he obtained diverse body photography's in movement on the same plate of glass: ten images per second with an exhibition of 1/1000 of second.

In 1892, Charles Émile Reynaud (1844-1918), the inventor of the "praxinoscope", he performed the first spectacle of cartoon animation with his optical theatre. The device was using the combination of two magic lanterns and images in movement that had been hand-painted, generally in colour. The resultant spectacle was known as "Pantomimes Lumineuses" and was realized in the Grévin Museum of Paris, where more than 12.800 representations were done to more than a half of million persons represented until it closed his doors in 1900. Three years before the invention of the cinematograph, already we can find cartoon animations [4].

At the beginning of the 20th Century, and already in "cinematographic age", the capture of the movement and the cartoon animation began their relationship with "rotoscope technology", which it was created by Max Fleischer (1883-1972).

At 90's, in the computer animation, the mocap technology is a procedure that the position is taken of a few points strategically placed in the body of a person and the positional information is transferred to a virtual avatar of the computer, with hierarchy skeleton. With this method, the virtual avatar transforms the position of the bones of the skeleton depending on the displacement of the capture, and with those points it is possible to move any movement of a person to a 3D character.

3. THE COMPUTER ANIMATION AND THE MOTION CAPTURE

Traditionally, in a film of computer animation, the artists construct the majority of characters in 3D with specific software. Later, the animators start transforming into some frames the variables articulated of the digitized models. Each of these variables describes a concrete movement. In this moment, the interpreters record his voices repeatedly to perfect his roles. When the voices are definitive, the animators begin the animation using the recording of the dialog of the interpreters as guide by the facial movement. The next process is to fit the first sketch of the film and to give to the surfaces genuineness with the textures, for example, the wrinkles of a face. Before the render (the generation of the final image), we must to apply the effects of lighting. With all the information, we can produce the finished images that are edited and are sent to postproduction for the final cut, generally in thirty five millimetres.

This process already is carried out by "Toy Story" (1995, John Lasseter, Pixar Animation Studios and Walt Disney Pictures), the first full-length film of the history created entirely by computer [5].

Nowadays, with mocap, the phase of animation of a character can be simplified enormously. With the technique of the keyframing and the blendshapes, the animation time increases a lot of time to create detailed expressions. On the other hand, with mocap, the work of animation of main characters is performed in a little time. In addition, the mocap contributes to the interpretation of the actors, the expressiveness of the bodies and the weight of the actions.

A film as "Total Recall" (1990, Paul Verhoeven, Sony Pictures and Columbia), already uses the technology of the mocap in some scene. In "Final Fantasy: The Spirits Within" (2000, Hironobu Sakaguchi, Square, Sony Pictures and Columbia) it is usual. In "The Polar Express" (2004, Robert Zemeckis, Warner Bros.), the facial mocap was used for first time. In this film, were captured the expressions of the North American actor Tom Hanks to interpret a 3D character with a similar look and gestural very notably. In "Monster House" (2006, Gil Kenan, Sony Pictures and Columbia) Mark Sagar introduces *Facial Acting Coding* (Facs) and the mocap.

Nowadays, the mocap is working with optical systems, like Vicom, with more than 48 cameras in some cases, and also we

work with a mixed system with mocap and keyframing. The producer Weta Digital is a good example: One of their works is Gollum from "The Lord's trilogy of the Rings" (2001, 2002, 2003, Peter Jackson) or King Kong (2006, Peter Jackson). In these films, to the British actor Andy Serkis, who was interpreting to the beast, his movements were captured, both of the body and of the face, by mocap, and later the animators were basing on this interpretation to animate directly the gestures of the beast. For the film, only there was the facial capture and they applied this capture to the animation with a rotoscope technique. The animators had the gorilla in 3D model with different controls (Motion Editing) and thanks to the capture, they could have a very near idea of what they wanted to animate.

4. EVALUATION OF EMOTION

In the field of the psychology there is not a consensus on describing the meaning of emotion [21], because there are a lot of definitions. But in the last years, there has been a general consensus that the emotions can differ and can be measured [1, 19]. There are 3D models to represent and measure the emotions (Figure 1).

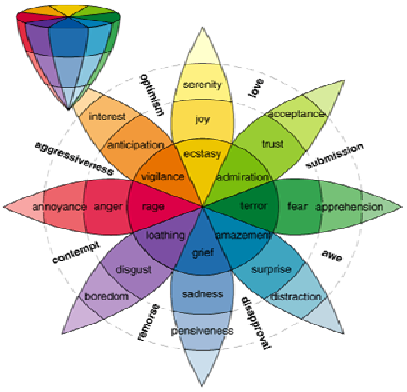


Figure 1.- The nature of Emotions [19]

Also we can find 2D simplified representations to measure the emotions. Two of them are: "Feeltrace", developed at Queen's University of Belfast and the most popular and recognized system: IAPS (International Affective Picture System [13]), revised and replied in several studies to check out its validity within diverse cultural frameworks [2, 20, 21].

The system studies the reaction of the user in front of several images including different semantic categories. The IAPS system is defined as an effective method to check out abnormal behavior and emotional dysfunctions in several types of users [8, 12, 14], and it quantifies the emotions at a two-dimensional representation level: the level of Valence (with a minimum value of 1 and a maximum of 9) and the Arousal (or activation, level of nervousness that an image exerts to a user, with values ranging from 1 (calm) to 9 (maximum excitement):

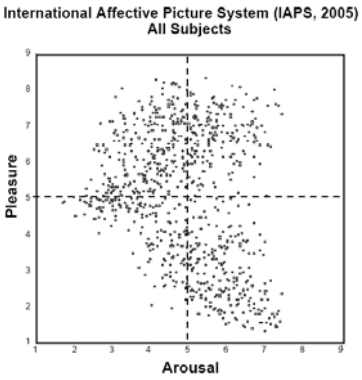


Figure 2.- 2D emotional representation. IAPS [29]

We can find works centered on evaluating the differences of perception for the image depending on the type of user as well as depending on the type and amount of emotions that the image provokes in the mentioned viewer [11]. Like we can read in the mentioned references, the user assigns different descriptors it depending on his origin and education or his personal influences [7].

In conclusion, this system is recognized like a valid method to measure all kinds of emotional answers of a user in front of a visual stimulus. From this affirmation, and adapting the levels and values that we want to study, we will be able to use it as a measuring method and an analysis in our investigation.

5. METHODOLOGY

To evaluate our hypothesis, we have made a specific test based on SAM (Self Assessment Manikin), used to evaluate the emotional answer of the user in the method IAPS [13].

We will show to the users 1 minute fragments from 6 animation movies, according to the type of facial animation that has been implemented. The fragments will show on a projector screen with a 1,9m diagonal, a resolution of 1024x768 and in a classroom with controlled lighting and "ideal distance" and angle of visualization [7].

We have worked with a group of undergraduate students (n = 30) experts in computer animation and the film clips have been taken from:

- Toy Story (Pixar/Disney, 1995): Facial animation of Andy. Keyframe Facial Animation with cartoon stylish.
- Final Fantasy (Square/Columbia, 2001): Facial animation of Aki. Keyframe Facial Animation with real aspect.
- The Incredibles (Pixar/Disney, 2004): Facial animation of Bob. Complex Keyframe/Blenshapes Facial Animation with cartoon stylish.
- Monster House (Columbia, 2004): Facial animation of DJ. Mocap Facial Animation with cartoon stylish.

- Polar Express (Warner Bros, 2004): Facial animation of Hero boy. Mocap Facial Animation with real aspect.

- Beowulf (Warner Bros, 2007): Facial animation of Beowulf. Mocap Facial Animation with extreme real aspect.

The clips (see figure 1) were selected from a representative scene of the film where the main character reflects emotional expression on his/her face. All the films and main characters were made in 3D computer animation with geometry, textures and lighting.

Figure 1.- Screenshot of the selected animation films

To evaluate the three main levels previously cited (sympathy, narrative involvement and narrative realism), we have included the following questions to measure them:

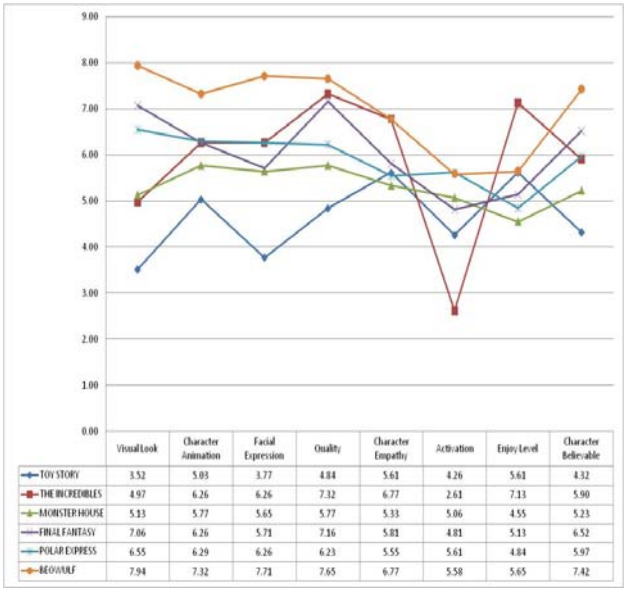
1. Narrative Realism:
    a. Evaluate the "visual look" of the scene and the character (from 1, non realistic to 9, very realistic).
    b. Evaluate the "animation" of the character (from 1, non human movement to 9, very realistic human movement).
    c. Evaluate the "facial expression" of the character (From 1, non realistic to 9, very realistic).
    d. Evaluate the technical quality of the scene (modeling, texture, lighting, rendering) (From 1, low quality to 9, high quality).

2. Sympathy:
    a. Evaluate the empathy or identification level with the character. In other words, are you happy or unhappy watching this type of character? (From 1, unhappy to 9, very happy).
    b. Evaluate the level of arousal such as excitement level that the character transmits us.

3. Narrative Involvement:
    a. How much did you enjoy watching this scene? (From 1, nothing to 9, a lot of).
    b. Was the character believable into the scene? (From 1, nothing to 9, a lot of).

6. RESULTS AND CONCLUTIONS

We can see in Figure 2, the results of the test that affirm that first hypothesis is not true. It does not have a correlation between the improvement that the mocap offers and the supposed improvement in the empathy of the characters.



	Visual Look	Character Animation	Facial Expression	Quality	Character Empathy	Activation	Enjoy Level	Character Believable
TOY STORY	3.52	5.03	3.77	4.84	5.61	4.26	5.61	4.32
THE INCREDIBLES	4.97	6.26	6.26	7.32	6.77	2.61	7.33	5.90
MONSTER HOUSE	5.33	5.77	5.65	5.77	5.33	5.06	4.55	5.23
FINAL FANTASY	7.06	6.26	5.71	7.16	5.81	4.81	5.13	6.52
POLAR EXPRESS	6.55	6.29	6.26	6.23	5.55	5.61	4.84	5.97
BEOWULF	7.94	7.32	7.71	7.65	6.77	5.58	5.65	7.42

Figure 2.- SAM test Median values

Polar Express, Monster House and Beowulf are the films that use the motion capture. Only Beowulf obtains high value of empathy. Other film with high empathy is The Incredibles (Figure 3).

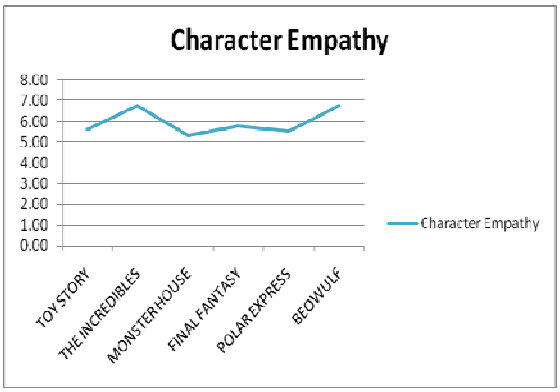


Figure 3.- Character Empathy Median values

We notice that a cartoon animation as The Incredibles has been evaluated with high score in realism because of the quality of the film. Viewing the "enjoy level", we can estimate that the contribution of the mocap does not improve the empathy level. Films like The Incredibles or Final Fantasy improve the results in facial expression in front of Polar Express or Monster House realized with mocap.

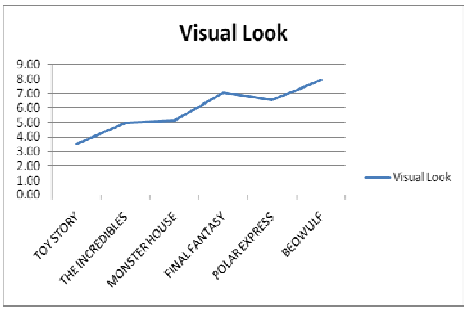


Figure 4.- Visual Look Median values

Beowulf, Polar Express and Final Fantasy have a high level of activation but with low level of enjoy. These results explain that it is very difficult for an animator to achieve realistic results, able to overcome the expectations of human observers who are experts at watching faces. It is easy, that results to fall in the "creepy appearance category" known as the “Uncanny Valley” (Figure 5), [15].

The “Uncanny Valley” hypothesis was introduced in a study of the emotional response of humans to robots and other non-human entities: as the appearance becomes more humanlike the response of a human is increasingly positive and empathic, until a certain point where the response quickly becomes strongly repulsive; then, as the appearance and motion of the entity evolves to be (almost) undistinguishable from a human being, the emotional response and familiarity approaches human to human empathy levels. This gap of repulsive response provoked by an “almost human” entity is called the “Uncanny Valley”. (Original graph by Dr. Mori 1982).

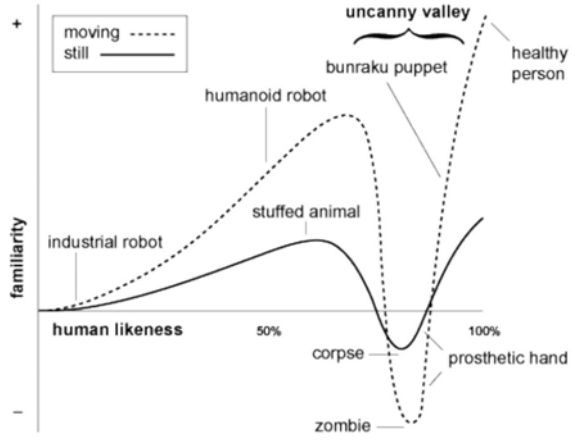


Figure 5.- The “Uncanny Valley”

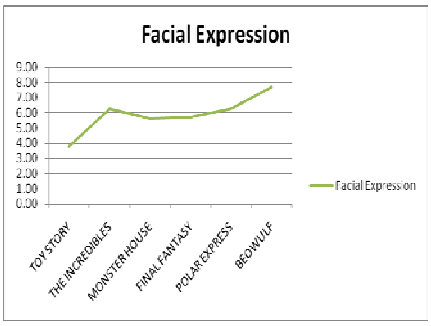


Figure 6.- Facial Expression Median values

We can conclude that the mocap is a good tool for production but the character developed with mocap has not more empathy than others methodologies.

It is necessary to create avatars with an image and real expression to jump the uncanny valley. But we need a lot of computational power. If it was not possible to reach this realism, the results conclude that we can reach high empathy levels with not real characterization of the character (cartoon), without need for an animation with mocap. If we take a look at the graph in figure 3, the most communicative face is the visual look and the animation of The Incredibles and not the most realistic like Beowulf.

The second conclusion is the no influence for communicative purposes of the mocap technology. Mocap is important for production but no better communicative results for mocap than keyframe or blendshapes animation.

The films that use the mocap technology are Monster House, Polar Express and Beowulf, and the only one that uses mocap and create character empathy is Beowulf.

We can create perfect faces with new engines that jump the valley and reach to create believable faces. In this path, our aim is to investigate new technologies like Universal Capture, Imagemetrics, Contour (Mova) or Aguru, that are new mocap technologies that maybe overpass the uncanny valley.

7. REFERENCES

[1] BOEHNER, K., DePaula, R., Dourish, P., Sengers, P. (2007). How emotion is made and measured. International Journal Human-Computer Studies , 65 (4), 275-291.

[2] BRUNO Verschuere, Geert Crombez, Ernst Koster. (2007). Cross Cultural Validation of the IAPS. Ghent, Belgium: Ghent University.

[3] CURRIE, G. (1997). The paradox of caring. Fiction and the philosophy of mind. In M. Hjort & S. Laver (Eds.), Emotion and the arts (pp. 63-77). New York, Oxford: Oxford University Press.

[4] DURAN, Jaume, El cinema d’animació nord-americà, Barcelona: UOC, 2008.

[5] DURAN, Jaume, Guía para ver y analizar Toy Story, Valencia - Barcelona: Nau Llibres - Octaedro, 2008.
[6] FISKE, S. T., & TAYLOR, S. E. (1991). Social Cognition. New York: McGraw-Hill.
[7] FONSECA, D., García, O., Duran, J., et al. (2008) *An image-centred "search and indexation system" based in user's data and perceived emotion*. 3rd ACM Int. Workshop on Human-Centered Computing. Vancouver, Canada, Pag. 27-34. http://doi.acm.org/10.1145/1462027.1462032
[8] FRANCISCO Aguilar, Antonio Verdejo, Mª Isabel Peralta, María Sánchez, Miguel Pérez. (2005). Experience of emotions in substance abusers exposed to images containing neutral, positive, and negative affective stimuli. (Elsevier, Ed.) Drug and Alcohol Dependence (78), 159-167.
[9] GENDLER, T. S., & KOVAKOVICH, K. (2005). Genuine rational fictional emotions. In M. Kieran (Ed.), Contemporary Debates in Aesthetics and the Philosophy of Art (pp. 241-253): Blackwell Publishing.
[10] GRAESSER, A. C., OLDE, B., & KLETTKE, B. (2002). How does the mind construct and represent stories? In M. C. Green, J. J. Strange & T. C. Brock (Eds.), Narrative Impact: Social and Cognitive Foundations (pp. 229-262). Mahwah, NJ: Lawrence Erlbaum.
[11] HONG, S., Ahn, C., Nah, Y., Choi, L. (2005). Searching Colour Images by Emotional Concepts. Searching Colour Images by Emotional Concepts, 3597/2005.
[12] HOUTVEEN JH, Rietveld S, Schoutrop M, Spiering M, Brosschot JF. (2001). A repressive coping style and affective, facial and physiological responses to looking at emotional pictures. International Journal of Psychophysiology, 42, 265-277.
[13] LANG, P., Margaret M. Bradley & Bruce N. Cuthbert. (2005). International Affective Picture System: Affective ratings of pictures and instruction manual. University of Florida, NIME. Gainesville: University of Florida.
[14] MIKELS, Fredrickson, Larkin, Lindberg, Maglio, Reuter-Lorenz. (2005). Emotional category data on images from the IAPS. Behaviour Research Methods, 37 (4), 626-630.
[15] MORI, M. The Uncanny Valley (1970). Energy 7, 4, 33-35
[16] NABI, R. L., Stitt, C. R., Halford, J., & Finnerty, K. (2006). Emotional and cognitive predictors of the enjoyment of reality-based and fictional television programming: An elaboration of the uses and gratifications perspective. Media Psychology, 8(4), 421-447.
[17] OATLEY, K. (2002). Emotions and the story worlds of fiction. In M. C. Green, J. J. Strange & T. C. Brock (Eds.), Narrative Impact: Social and Cognitive Foundations (pp. 39-69). Mahwah, NJ: Lawrence Erlbaum.
[18] OLIVER, M. B. (2003). Mood management and selective exposure. In J. Bryant, D. Roskos-Ewoldsen & J. Cantor (Eds.), Communication and Emotion: Essays in Honor of Dolf Zillmann (pp. 85-106). Mahwah, N.J.: LEA.
[19] PLUTCHIK, R. (2001). The Nature of Emotions. (H. a. Row, Ed.) The Nature of Emotions, 89 (4), 344-350.
[20] RAMIREZ O., Molto, J. et Al. (1999). Un Nuevo método para el estudio experimental de las emociones: El international affective picture system (IAPS). Adaptación Española. Revista de Psicología General y Aplicada / Journal of Psychophysiology, 55/22 (1), 55-87/312-313.
[21] RAQUEL Chayo, A. E. (2003). Valencia, activación, dominancia y contenido moral, ante estímulos visuales con contenido emocional y moral: Estudio en población mexicana. Revista Española de Neuropsicología, 3-4 (5), 213-225.
[22] SHAPIRO, M. A., & CHOCK, T. M. (2003). Psychological processes in perceiving reality. Media Psychology, 5(2), 163-198.
[23] YANAL, R. J. (1999). Paradoxes of emotion and fiction. University Park, PA: The Pennsylvania State University Press.
[24] ZILLMANN, D. (1994). Mechanisms of emotional involvement with drama. Poetics, 23, 33-51.
[25] ZILLMANN, D. (1988). Mood management through communication choices. American Behavioral Scientist, 31, 327-340.