**KEER2010, PARIS | MARCH 2-4 2010**
**INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH 2010**

# EMOTIONAL INTERACTIVE STORYTELLER SYSTEM

**Peter BLEACKLEY[a], Sylwia HYNIEWSKA[b], Radoslaw NIEWIADOMSKI[b], Catherine PELACHAUD[b] and Marc PRICE[a]**

[a] *BBC Research and Development, United Kingdom*
[b] *LTCI-CNRS, Telecom ParisTech, France*

## ABSTRACT

This paper presents a concept and initial evaluation of the emotional interactive storyteller system. Our concept is derived from our knowledge of the pre-medieval oral storytelling tradition and its application to therapeutic storytelling. Our initial evaluations are based on a mock-up of the concept where emotional content of the story is reinforced with an expressive agent, which was subjectively tested with a sample of volunteer users. The results of the evaluation are very helpful, confirming that the concept is useful and meaningful in general to the user, and providing us with valuable feedback in shaping the ongoing design for a full implementation.

Keywords: **storytelling, embodied conversational agent, emotional behaviour**

## 1. INTRODUCTION

The birth of print marked the beginning of a paradigm shift in storytelling. Prior to this shift, a storytelling was an event conveyed by word of mouth in front of a live audience. The large body of studies in the oral storytelling traditions (e.g. Campbell[1], Lord[2], Propp[3]), inform us that repertoires consisted of stories having a similar overall structure, drawing on common sub-plots and narratives. Yet no story was told the same way twice. There is much conjecture about why this was so, and we are interested in the hypotheses used in the field of therapeutic metaphor [4].

In essence, these hypotheses state that the **meaning** of a story exists at a higher level of abstraction than the **plot**. Thus, by rearranging sub-plots, altering or improvising the details of events, etc, the same story told in different ways can have very different meanings (metaphor). A core element of this is the sequence of affective states through which the audience is carried. In live storytelling, the storyteller uses his highly honed skills in sensory acuity to continuously read the affective state of the audience. He can thus 'pace' the audience [4], and improvise accordingly to 'lead' the audience through the desired sequence of affective states to convey the underlying message.

In print and all other broadcast media, the feedback loop is removed because there are no means for sensory acuity and improvisation. These media only permit non-interactive, or 'linear', storytelling. However, with modern interactive entertainment technology, we have the means for improvisation. The missing link is the technology that has the sensory acuity to recognise the audience's changing moods in real-time.

We are participating in the CALLAS Project, a FP6 integrated project that promises to provide this missing link. CALLAS is a research project with a leading-edge consortium of eighteen companies and research institutions. It is delivering multimodal affective and reactive interfaces for Digital Media, based on various elements of speech and gesture. In this paper, we present our initial evaluation of our approach to emotional interactive storytelling within CALLAS.

The remainder of this paper is organised as follows. In the next section we explore the current state of the art in technologies for emotional storytelling. Section 3 gives a description of the general concept of a storyteller application. In Section 4 we present the results of our mock-up evaluations. We conclude in Section 5.

## 2. STATE OF ART

Expressions of emotions play an important role in this human-like, multimodal communication. Several results demonstrate that the application of facial expressions in embodied agents is justified. The software is perceived as more engaging when it uses emotional expressions [5]. The emotional expressions were found to be very useful in understanding the (ambiguous) text messages [6]. But the agent does not only need to express some facial expressions, it should display facial expressions that are adequate to the situation. Inattentive application of facial expressions leads to the effects opposite to the expected [7][5]. The facial expressions influence also the perception of an agent. Indeed, the same agent with different facial expressions is perceived as more credible and trustworthy [8]. Showing an emotional expression increases user's attention. People spend more time interacting with an agent that expresses emotions than with one displaying a neutral expression [5].

On the other hand the concept of emotional storyteller has rarely been studied. Notably in [9], Blom and Beckhaus present their concept of Emotional Storytelling in Interactive

Storytelling systems, which is very similar to the concept presented in this paper. Blom and Beckhaus' concept differs in that they do not consider the possibility of pacing the audience's current emotional state and leading them to the state desired by the author. They also did not make any evaluations of their concept. In [10] an embodied conversational agent was used as a co-spectator which influences the human's emotional reactions for the presented content through its emotional displays.

## 3. CONCEPT

### 3.1. Motivation and aims

As this is an entirely new application, we have a fairly blank canvas and we need to determine what will work best for the audience: what kind of interactions; the look and feel; the motivators for our audiences to use it; etc. Moreover, we need to verify our thesis as outlined above.

To do this, we need to produce and test an application that behaves like a live storyteller, in the sense of leading the audience through a desired sequence of affective states over the course of a story. In this paper, our aim was to evaluate our overall concept with some user tests.

### 3.2. Application

We would ultimately like to apply this concept to all storytelling applications within the broadcast domain. For now, we focus on current affairs, which takes us way beyond the formal story structures of the oral traditions.

The application consists of an ECA (embodied conversational agent) coupled with an image display. The ECA has the role of a storyteller and the image display shows images that relate to the story content. For this purpose we use Greta agent [11]. The story is broken down into scenes, with an image associated to each scene.

When the application begins, the user is presented with the image for the first scene. The ECA invites the user to say some words about the scene. When the user has finished speaking, the ECA conveys the story of the scene to the user. It also uses emotional displays. At the beginning of the next scene, the next image is presented. Once again, the ECA invites the user to say some words about the scene. This turn-based interaction continues through to the end of the story. The Greta agent uses facial and body behaviours to share an emotional state with the user. Its emotional behaviour model is based on a theoretical ground [12] [13] and driven by a precise analysis of natural audio-visual data [14]. This model that generates emotional behaviour as sequential and multimodal is described in [15].

The application is essentially trying to lead the user through the exact sequence of affective states for the story as prescribed by the story author. Through the system, the author of the content may influence the implication and emotional reactions of the audience, like a movie director or music composer. The originality of this type of application is its use of an interactive and expressive storyteller, which can be combined with music to induce affective states in the user.

Under the hood, the application analyses the user's speech and gesture at each juncture in the story, to determine their affective state at that instant. The measured state vector is compared to a predefined ideal state vector for that point in the story.

The result of the comparison is then used to vary the way in which the next scene is conveyed to the user by the ECA, in a bid to bring their state closer to the next ideal state in the story. For example, if the user is not frightened enough, the next scene will be gradually expressed in a more frightening way than the norm.

### 3.3. Description of Mock-up

The aim of the mock-up is to look and feel to the user like the full application as we had originally conceived it. The mock-up presents the user with an ECA, an image display with a few words of text, and some music (see Figure 1). The user interaction is 'turn-based', speaking when invited to by the ECA, and moving through the scenes using the spacebar. The mock-up application makes no use of the user's spoken inputs, and the presented story is linear.

The story content for the mock-up is based on the Sichuan earthquake disaster of May 2008. We make use of six images, each associated with two sentences of narrative, to briefly convey the impact of the disaster to people's lives in the region. We use this subject because it is topical and has strong affective reactions associated with it.

## 4. USER TESTING

A sample of 27 volunteers was recruited from the staff of a major broadcaster. Each volunteer was asked to fill in a questionnaire asking for demographic information, before starting the application. Participants were asked to report their internal state by means of the Self Assessment Manikin [16] at each point where they were asked for a response, and one more time at the end of the experiment. The Self Assessment Manikin (SAM) test is a commonly used experimental procedure [16]. It is a pictorial questionnaire enabling participants to assess their internal state on three dimensions, namely Pleasure (P), Arousal (A) and Dominance (D) (also referred to as PAD) [17]. Its advantage is that it is a cross-cultural and language independent test [18]. It is also text-free, therefore avoiding possible bias linked to a linguistic description of emotions [19]. After using the demonstrator, they were asked to fill in one further questionnaire concerning their opinions of the experience. Volunteers were not told at any point that the demonstration was a mock-up.

Two configurations of the system were tested. These were voice only, and voice and ECA versions of the storyteller.

### 4.1. Procedure

In the experiment we use two different conditions. In the first condition (C1) the ECA is not used; the system displays pictures and produces voice prompts and comments (the control condition). In the second condition an ECA is displayed in addition. The display can be solely a face or a face completed with a whole body expression. Both displays are merged into a unique condition called embodied storyteller. It is referred as C2 condition. The ECA displays

emotional expressions that correspond to the presented content, prompts the user to comment the pictures and says its comments about each picture.

**Figure 1:** User Testing of the Interactive Storyteller.

Our aim is to verify if the user's emotional state is changed during the story telling. Thus, we asked subjects to report their emotions before, during and at the end of each story telling session. In more details, before the beginning of the emotion induction procedure (i.e. in time $t_0$) participants were asked to rate their internal state using PAD. This evaluation was repeated three times (after each of the three parts of the story, $t_1$ - $t_3$). The participants reported their states on a 9 point PAD scale, with each axis running from -1 to +1 [6].

After the experiment, they were asked to report the answers to the following questions on a five-point scale running from -1 to +1:

- Q1) How much did you like the demonstration?
- Q2) Did the storyteller help in conveying the content of the story?
- Q3) Did the storyteller help you engage in the story?
- Q4) Would you use a system like this again in the future?
- Q5) How appropriate were the storyteller's reactions?
- Q6) Did the storyteller help you to perceive the emotional side of the story?
- Q7) How involved did you feel in the story?

They were also asked to give free comments about the system.

### 4.2. Hypotheses

An overall effect of the story on the attributions of the three dimensions of P, A and D was expected (C1 and C2 confounded). Specifically:

- (H1a) The level of reported Pleasure (P) is expected to diminish in reaction to the sad story. The P value will be lower after the emotional induction, i.e. P0 < P1, P2, P3 where *Pi* is a value of pleasure in time $t_i$, *i=0..3*,
- (H1b) The level of reported Arousal (A) will be altered during the story. The A value will be affected by each presented extract. The A value will be different at $t_0$ to the later stages, i.e.: AO ≠ A1, A2, A3, where *Ai* is a value of arousal in time $t_i$, *i=0..3*,
- (H1c) The presentations leading to an appraisal of lack of possible control over the situation will decrease the reported feeling of dominance (D). The D value will decrease after the induction, i.e. D0 > D1, D2, D3, where *Di* is a value of dominance in time $t_i$, *i=0..3.*

We also expected (H2) a stronger effect of an embodied storyteller (condition C2) in comparison with a voice only version (condition C1). Thus in this hypothesis we expected that in condition C2 the system would induce a greater change due to the multimodality than in condition C1 (voice).

Finally we expected (hypothesis H3a-g) positive answers to the questions Q1-Q7 i.e. that the users will appreciate the demonstration (H3a), the users will find the storyteller helpful in perceiving the content of the story (H3b), the users will find the storyteller helpful in becoming engaged in the story (H3c), users will like a system like this again in the future (H3d), the reactions of the storyteller will be evaluated as appropriate by the users (H3e), the users will find the storyteller helpful in perceiving the emotional side of the story (H3f) and the users will feel involved in the story (H3g).

### 4.3. Results

The sample consisted of 27 broadcasting company employees, of whom 16 were male and 11 were female. They ranged between 23 and 58 years of age, with a mean of 38 years and a standard deviation of 9 years.

All the participants but one described themselves as having spent most of their time in the United Kingdom over the past 5 years. Twenty-four stated their first language as English, and the three remaining as Urdu, Gujrati or Malay. Twenty-three volunteers described their ethnicity as white, two as Indian, and the remaining as one Pakistani, and one Other Asian (ethnicity categories were based on those used in the UK census). It will therefore not be possible to infer any effect of country of residence, first language or ethnicity from this sample.

12 participants saw the voice-only version of the storyteller, 15 the voice and ECA storyteller.

First we checked the emotional impact of the application on the user's reported state. A strong effect of the time in the story when the evaluation was collected was observed for the PAD scores (multivariate ANOVA, $p<.005$). The reported PAD values after each part of the experiment can be seen on Figure 2.

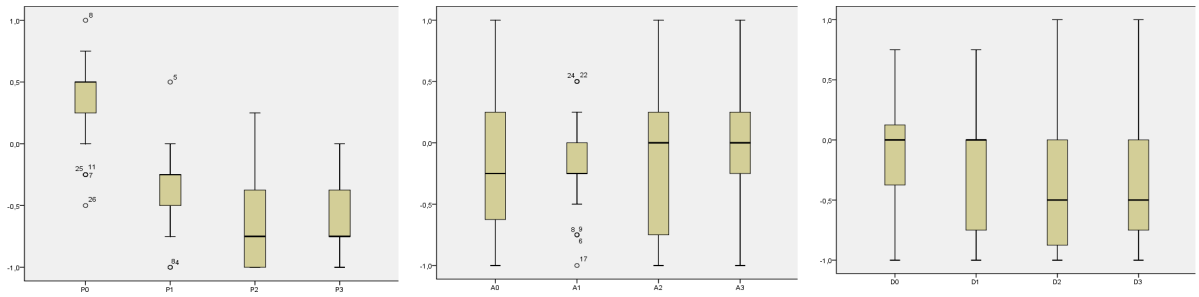


**Figure 2:** Reported PAD values in each segment of the story.

We realised pair wise comparisons between reported P, A, and D values. For pleasure, there was a difference between the preliminary evaluation and the three following ones: P0 (.36) and P1 (-.37), P0 and P2 (-.62), P0 and P3 (-.60) (Wilcoxon matched pairs test, p<.005 for a one tailed Exact sign). In the case of arousal, the difference between A0 (-.18) and each of the later reports (respectively -.15; -.10; -.06) was not significant (Wilcoxon matched pairs test, p>.05 for a one tailed Exact sign.). Finally, in the case of dominance the difference in reported values was significant between D0 (-.14) and D2 (-.37) (Wilcoxon matched pairs test, p>.005 for a one tailed Exact sign), and on the limit of significance for D0 and D3 (-.32) (p=.057) but not between D0 and D1 (-.22) (p>.05).

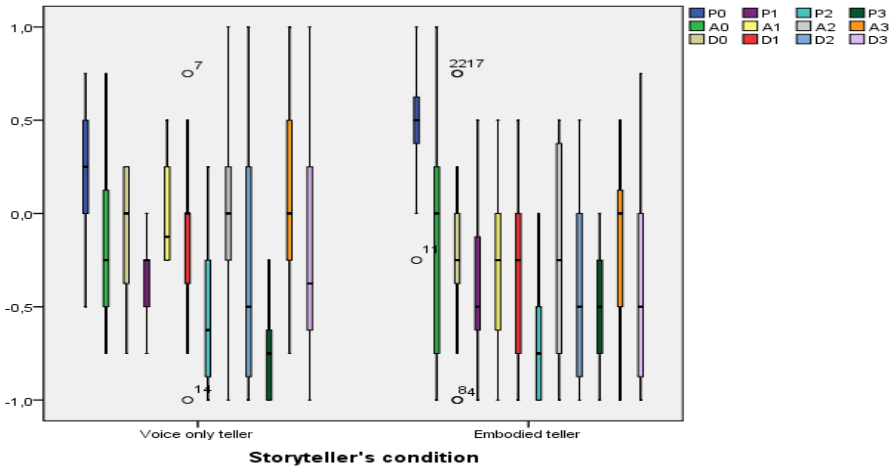


**Figure 3:** Effect of the storyteller's condition on the PAD values in the different times of evaluation.

The emotional induction effect between the two different storyteller conditions (C1-C2) of the experiment was also analysed, see Figure 3. The repeated measures ANOVA showed no effect of the application of the embodied conversational agent on the PAD values. Thus, whether the story was told by a voice only or an embodied storyteller the effect was similar.

In general participants liked more than disliked the demonstration (question Q1, 15 positive answers), and found the system helpful in conveying the story content (question Q2, 13 positive answers). They also found the storyteller's reactions appropriate (question Q5, 14 positive answers) and found it helpful in perceiving the emotional side of the story (question Q6, 13 positive answers). On the other hand our system only moderately helps in being involved in the

story (question Q3, 13 positive answers). People reported no strong engagement (question Q7, 10 positive answers). Finally our subjects were largely undecided about an eventual future use of similar systems (question Q4, ~25% positive, ~25% negative, and ~50% undecided answers).

In the open ended questions participants stated that: they liked the overall experience, finding it meaningful and useful. They liked the concept of having an empathic virtual storyteller. On the other hand, they found it also confusing to have the storyteller in a separate window to the image viewer. The implication being that it should instead be like a weather presenter, or indeed like any other kind of real (but informal) lecture-style presentation, with the presenter being in the same 'room' as the image viewer. What is more, they found the poor graphics quality of the storyteller off-putting. Most participants assumed that this is something that would be a necessary improvement if it were to become a product.. Concerning the study itself, people generally found the scoring set-up rather cumbersome, and it interfered with the flow of the demo.

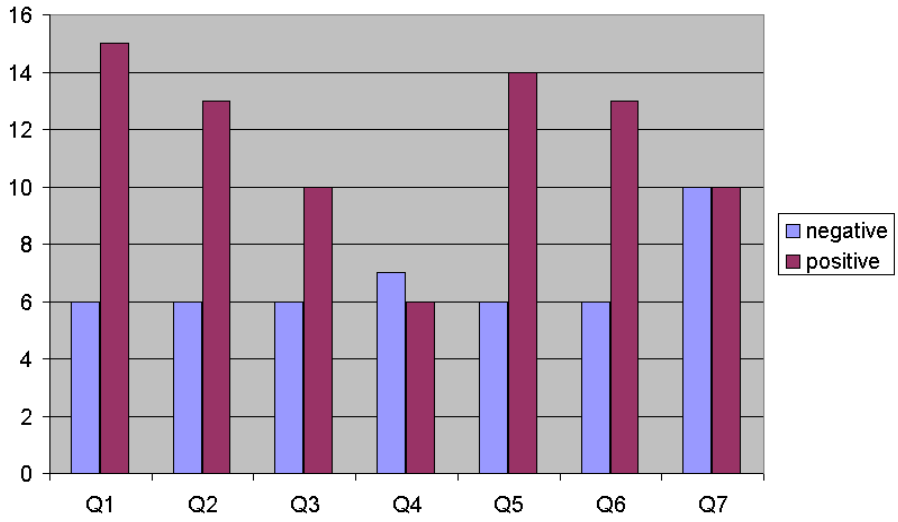


**Figure 4:** The number of positive vs. negative answers in questions Q1-Q7.

### 4.4. Discussion

Our aim in this experiment was to test the impact of a virtual storyteller as an emotion inductor. Two conditions were compared, a voice only news presenter and the same voice presenter with an embodied human appearance. The first hypothesis (H1) was verified. The storyteller system was an efficient emotion inductor system. The reported values of pleasure, arousal and dominance were significantly influenced by the experience.

Only the hypothesis H1b was not confirmed between H1a-c. These self assessments indicate that most of the participants were lead into more negative (H1a) and less dominant emotional states (H1c) than before the experience. At the same time reported arousal (H1b) did not change significantly. On the other hand the area of a PAD model that was often chosen by the subjects correspond to sadness (e.g. P = -0.5, A= -0.3, D= -0.4 in [8]). The mean reported PAD values may suggest that the system might evoke empathetic reaction in the users. However our questionnaire does not allow us to fully answer this question. The high individual variation was observed in user's answers even if most people followed the same trend (less valence, less

dominance). On the other hand studies have shown that the values for a specific emotional stimulus attributed by users may vary. This variation may be partly explained by personality traits. For example trait anxiety is associated to an attentional bias for threatful stimuli, such as angry faces (e.g [20]). Some studies show that adult attachment styles may influence the affective evaluation of social signals (e.g. [21]). Nevertheless, in our experiment the results along the P and D dimensions show sufficiently strong general tendency among the participants between the different sections of the experiment. Globally the level of pleasure and dominance diminished during the experiment and the tendency of the change is maintained between the different sections of the experiment. This effect should be studied more deeply in future work.

It is noteworthy that the stimuli used in the experiment are complex. They could be evaluated differently by the participants whose attention could be attracted to various elements of the presented scene. For example some may focus on the death of the inhabitants while others on weakness of the infrastructure and blaming the people involved in the construction.

The hypothesis H2 about the difference in emotional induction between agent based and voice only conditions was not confirmed. We did not find significant differences in reported emotional reactions between the embodied agent based (C2) and voice only (C1) application. This might be due to the poor quality of the animation, or the distraction caused by having to manually provide an affective scoring at each juncture, as reported in the open-ended questions. It could also be due to the ceiling effect as the story may not be expressed in more negative and arousing way then its content allows.

Moreover, we cannot ignore the choice of the population, with all participants being conversant with multimedia tools of high visual quality. This creates a bias, with a greater expectation of a high level standard of graphics. Indeed they reported only moderate engagement during the session with the system (H3g) and they were undecided about reusing the system in the future (H3d).

The remaining hypotheses (H3b, H3e, H3f) were generally confirmed. Participants found the system helpful in conveying the story content as well as perceiving the emotional side of the story. The reactions of the storyteller were appropriate.

## 5. CONCLUSION

In this paper, we have presented our emotional storyteller concept and our initial mock-up based evaluation of it. The results of the evaluation study are promising. The resulting user comments suggest design alterations are required to effect improvements in the quality of its implementation, rather than any fundamental flaws to the concept itself. Specifically, improvements to the graphics quality of the animated character, and the display layout are required. Our immediate future work is to test a full implementation of the concept, taking these suggested modifications into account.

## ACKNOWLEDGEMENT

Part of this research is supported by the EU FP6 Integrated Project CALLAS IP-CALLAS IST-034800.

## REFERENCES


[1] Campbell, J. *The Hero with a Thousand Faces*. New World Library, 2008.

[2] Lord, AB. *The Singer of Tales*. Harvard University Press, 1960.

[3] Propp, V. *Morphology of the Folk Tale*. University of Texas Press, 1958.

[4] Gordon, D. *Therapeutic Metaphors*. Meta Publications, 1978.

[5] Walker, JH, Sproull, L, Subramani, R. Using a human face in an interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*, Boston, Massachusetts, United States, pp. 85 - 91, 1994.

[6] Elliott, C. I picked up Catapia and other stories: A multimodal approach to expressivity for 'emotionally intelligent' agents. In *Proceedings of the First International Conference on Autonomous Agents*, Marina del Rey, CA, pp. 451-457, 1997.

[7] Becker, C, Prendinger, H, Ishizuka, M, Wachsmuth, I. Evaluating Affective Feedback of the 3D Agent Max in a Competitive Cards Game. In *First International Conference on Affective Computing & Intelligent Interaction*, Beijing, pp. 466-473, 2005.

[8] Rehm, M, André, E. Catch Me If You Can - Exploring Lying Agents in Social Settings. In *Int. Conf. on Autonomous Agents and Multiagent Systems*, Utrecht, pp. 937-944, 2005.

[9] Blom, K, Beckhaus, S. Emotional Storytelling. *IEEE Virtual Reality Conference*, Bonn, Germany, pp. 23-27, 2005.

[10] Charles, F, Lemercier, S, Vogt, T, Bee, N, Mancini, M, Urbain, J, Price, M, André, E, Pelachaud, C, Cavazza, M. Affective Interactive Narrative in the CALLAS Project, *Fourth International Conference on Virtual Storytelling*, Saint-Malo, 5-7 Dec 2007.

[11] Niewiadomski, R, Bevacqua, E, Mancini, M, Pelachaud, C. Greta: an interactive expressive ECA system, In *Proceedings of the Eighth International Conference on Autonomous Agents and Multiagent Systems*, Budapest, 2009.

[12] Shiota, MN, Campos, B, Keltner, D. The faces of positive emotion: Prototype displays of awe, amusement, and pride. *Annals of the New York Academy of Sciences*, 2003.

[13] Ekman, P, Friesen, W. *Unmasking the Face: A guide to recognizing emotions from facial clues*. Prentice-Hall, Inc., 1975.

[14] Niewiadomski, R, Hyniewska, S, Pelachaud, C. Evaluation of Multimodal Sequential Expressions of Emotions in ECA. In *Proceedings of Third International Conference on Affective Computing and Intelligent Interaction* (ACII 2009), 635-641. Amsterdam, 2009.

[15] Niewiadomski, R, Hyniewska, S, Pelachaud, C. Introducing multimodal sequential emotional expressions for virtual characters. *International Conference on Kansei Engineering and Research*, Paris, 2010.

[16] Bradley, MM, and Lang, PJ. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. Journal of Behavior Therapy and Experimental Psychiatry, 25(1), pp. 49-59, 1994.

[17] Mehrabian A. Analysis of the Big-five Personality Factors in Terms of the PAD Temperament Model. *Australian Journal of Psychology*, vol. 48, 2, pp. 86-92, 1996.

[18] Morris, JD. Observations: SAM: The self-assessment manikin, an efficient cross-cultural measurement of emotional response. *Journal of Advertising Research*, pp. 63-68, 1995.

[19] Grimm, M, Kroschel, K. Evaluation of natural emotions using self assessment manikins. In *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding*, Nov. 27-Dec. 1, pp. 381-385, 2005.

[20] Bradley, BP, Mogg, K, Falla, SJ, and Hamilton, LR. Attentional bias for threatening facial expressions in anxiety: Manipulation of stimulus duration. *Cognition and Emotion,* 12, pp. 737–753, 1998.

[21] Vrticka, P, Andersson, F, Grandjean, D, Sander, D, Vuilleumier, P. Individual Attachment Style Modulates Human Amygdala and Striatum Activation during Social Appraisal. *PLoS ONE*, 3(8), pp. 28-68, 2009.