Evaluating the impact of Embodied Conversational Agents (ECAs) attentional behaviors on user retention of cultural content in a simulated mobile environment

Ioannis Doumanis
Middlesex University
The Burroughs, Hendon
London NW4 4BT
+44 (0) 208411 6747
i.doumanis@mdx.ac.uk

Serengul Smith
Middlesex University
The Burroughs, Hendon
London NW4 4BT
+44 (0) 208411 6747
s.smith@mdx.ac.uk

ABSTRACT

The paper presents an evaluation study of the impact of an ECA's attentional behaviors using a custom research method that combines facial expression analysis, eye-tracking and a retention test. The method provides additional channels to EEG-based methods (e.g., [8]) for the study of user attention and emotions. In order to validate the proposed approach, two tour guide applications were created with an embodied conversational agent (ECA) that presents cultural content about a real-tourist attraction. The agent simulates two attention-grabbing mechanisms - humorous and serious to attract the users' attention. A formal study was conducted to compare two tour guide applications in the lab. The data collected from the facial expression analysis and eye-tracking helped to explain particularly good and bad performances in retention tests. In terms of the study results, strong quantitative and qualitative evidence was found that an ECA should not attract more attention to itself than necessary, to avoid becoming a distraction from the flow of the content. It was also found that the ECA had an inverse effect on the retention performance of participants with different gender and their use on computer interfaces is not a good idea for elderly users.

Categories and Subject Descriptors

H.5.2: **[Information Interfaces and Presentation]**: User Interfaces – *Evaluation/methodology, Prototyping, User-centered design, Interaction styles (e.g., commands, menus, forms, direct manipulation).*

General Terms

Design, Experimentation

Keywords

Embodied Conversational Agents (ECAs), Eye Tracking

1. INTRODUCTION

To date, little research in the ECA community has been conducted using advanced techniques for usability research like eye tracking, let alone, using a technique that combines data from eye tracking, with data from facial expression capturing. The human face is one

of the strongest indicators of a human's cognitive state and hence how humans perceive stimuli (information, images, etc.) [15]. A technique that combines data from facial expression recognition and eye-tracking can augment any traditional techniques for accessibility evaluation. In one study [1] eye-tracking was used to measure the amount of attention drawn to ECAs. Participants interacted with an ECA that informed them about different sorts of wine through both verbal and textual means. It was shown that an ECA attracts and holds the participants' visual attention in the interface, but also that this focus may have distracted from the products about which the agent was providing information. However, it could be argued that since participants devoted most of their time to reading the text appearing in a speech bubble above the agent it was hard for them to keep up with the agent's speech discourse. A similar study [2] on the attention-capturing quality of agents' utilizing gaze tracking reached the same conclusion. In particular, it was found that a character agent captures attention to a greater extent, than an animated arrow and that the participants in the latter condition needed more time to react than in the former. This extra time was interpreted as lack of concentration on the game by the paper's authors. One could argue, though, that the longer response times in the character agent condition may be the result of the additional cues (facial expressions, etc.) that participants had to account in making their decisions. In this way, the greater complexity of the situation may have led to increased response times. Another study [3] investigated the use of eye gaze as input to an agent-based virtual sales scenario. The system adapts the presentation to match the user's interests, and reacts appropriately if the user is inattentive. In an exploratory study, two versions of the system were compared: a fully interactive version which analyzed the user gaze behavior in real-time and provided appropriate reactions to interests/disinterest, and a pseudo-interactive version which was based on randomly assigned interruptions. They found that in the interactive version, the agents guided the user's attention more successfully to the content of the presentation than in the pseudo-interactive version. In this study, the proposed approach aims to investigate whether an ECA using an attention grabbing mechanism enhances the participants' ability to retain information from presentations in an archaeological attraction or not.

2. EXPERIMENTAL DESIGN

This experiment was carried out in collaboration with eMarketView [4] company which provided the lab, equipment and test subjects and gave us the ability to moderate the sessions

remotely. The company uses the EyeTechDS eye tracker[1] and EyeTellect's gaze tracking software [11] as the standard tools for their user experience research projects. The lab setting was as follows: Two tour guide applications featuring two Embodied Conversations Agents (ECAs) were run on a desktop computer with Microsoft Windows 7 and resolution of 1920 x 1080 pixels. The ECA systems were developed in Microsoft Visual Studio 2010 [9] using the Haptek avatar engine [10]. The layout of the system was designed to simulate a tablet mobile device (see Figure 2). The first ECA employs a range of serious or humorous strategies to grab the attention of the participants to presentations about four attractions in the archaeological attraction. An example of a humorous message is, *"I guess the description of churches is a boring topic for most people, but please bear with me. Perhaps one day you will play in the TV show who wants to be a millionaire and win thanks to me"*. The serious messages were more intrusive and sometimes rude. An example is, *"Please pay attention to me and try to keep quiet. I have only a limited amount of time to complete the tour"*. The second system without any attention-grabbing mechanisms has provided information about four locations of the castle. Each presentation was designed to evoke at least some content-related emotions such as happiness. We hypothesized that the participant's facial expressions in each presentation would indicate his/her underlying emotional state. Facial expressions were recorded using a standard web camera attached to the computer. The videos were recorded at 25 frames per second and with a resolution of 320 x 240 pixels. Furthermore, eye tracking was used to identify the section of the interface, where the participant was looking at when the particular expression occurred.

Table 1. The experimental design

Participants	F/M	Attention Grabbing		Non-Attention Grabbing
		Content Complexity	Feature Type	Content Complexity
1	F	Simple	Serious	Complex
2	F	Simple	Serious	Complex
3	F	Simple	Serious	Complex
4	M	Simple	Humorous	Complex
5	M	Simple	Humorous	Complex
6	M	Simple	Humorous	Complex
7	M	Complex	Serious	Simple
8	M	Complex	Serious	Simple
9	M	Complex	Serious	Simple
10	F	Complex	Humorous	Simple
11	F	Complex	Humorous	Simple
12	F	Complex	Humorous	Simple

We measured the type of ECA (attention-grabbing vs. non-attention grabbing), type of content (simple vs. technical) and order of presentation (simple then complex vs. vice versa) as the between-subject variables. The attention-grabbing strategy used by the ECA (humorous vs. serious), the type of content (simple vs. complex) and the participants' gender (females then males vs. vice versa) were manipulated as within-subjects variables (see Table 1). Participants were randomly assigned to the eight experimental conditions: 1) Serious AG ECA with the simple content vs. non-AG ECA with the complex content or 2) Humorous AG ECA with the simple content vs. non-AG ECA with the complex content or 3) Serious AG ECA with the complex content vs. non-AG ECA with the simple content or 4) Humorous AG ECA with the complex content vs. non-AG ECA with the simple content.

---

[1] The eye tracker's full specifications can be found at http://www.eyetechds.com/

2.1 Participants

Thirteen participants (7 women and 6 men age between 35 and 65 with a normal cognitive ageing without age associated cognitive decline) were recruited by eMarketView for the experiment. In order to test the experimental settings, a short trial has been carried with one of the female participants. The remaining twelve participants were assigned equally to the experimental conditions at random. The participants had varying mobile computing experience and educational backgrounds.

2.2 Task

The experiment began by asking the participants to read the experiment brief and followed by answering any questions they might have. Then, they were asked to go through four short presentations in turn about the castle of Monemvasia. The presentations varied in terms of complexity. The technical content included information specific to the architecture of the four locations, while the simple information of general interest. After participants had listened to the presentations for all four locations, they were provided with a list of randomized keywords and were asked to fill-in a retention test on the information they heard during the presentations. The list of keywords was provided to help participants in recalling Greek names.

2.3 Measures and methods

2.3.1 Post-task Questionnaires and Retention Test

The object recognition (yes/no) questionnaires were used to indicate whether they saw specific objects during the experiment. The retention performance of the participants was collected through fill-in-the-blanks tests where participants had to fill-in words missing from sentences, the ECA uttered during the presentations. The participants were also asked to rate the confidence of their answer on a ten-point scale (1 = completely at random, 5 = not so confident, 10 = totally confident).

2.3.2 Face Recording

A standard web camera attached to a desktop computer recorded the participant's face from a straight angle. Because of budget and time constraints, the video files of each presentation were analyzed by one expert, specializing in the field of cognitive psychology. The expert followed a two stage approach to analyse the video files.
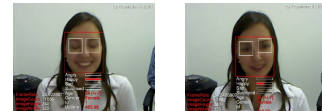

Figure 1. Detection of participants' facial expressions

In particular, the expert first manually analyzed the video files using a map of emotions and facial expressions [5] and then, used the FaceDetect demo application [6] to validate the findings. The application uses the SHORE engine that outputs four distinct facial expressions: happy, sad, angry and surprised with a visual indication of the intensity of the expression. Although SHORE can also output numerical values (in a range of 0 – 100 with a higher value, meaning a more intense expression) for the intensity of an expression, the Face Detect demo application does not offer this feature. This functionality requires embedding SHORE into a custom software application through its API. Given the project time and budget constraints, we decided to instruct the expert to qualitatively validate the findings of the manual analyses of the video files.

2.3.3 Eye Tracking: Gaze Trails and Heat Maps

Gaze trails (see Figure 4) provided data on the sections of the interface on which the participants cast their eyes, in which order and for how long. Heat maps (see Figure 3) provided an amalgamation of where each participant looked at and for how long. These measures were analyzed and the results were correlated with the data collected from the retention tests.

3. RESULTS AND DISCUSSION

Despite the small-sized subject pool (12 participants) we decided to use statistical analyses, because we wanted to make better sense of the gathered data, that is, to present our confidence in the numbers. Although, we recognize that a small sample size means that any significant results are less than "statistically valid", it does not mean that they are "less than valid". In fact, the applied user research literature [12] [13] indicates that small sample sizes generate viable results when it comes to large differences between designs or to discovering common usability problems. Therefore, the generalizability and the depth of the results reported below should be considered with caution, as further research is needed with a bigger-sized subject pool.

3.1 Object Recognition Questionnaires

Table 2 shows the total number of objects participants confirmed in the questionnaire under each condition. Overall, participants were able to recognize the objects/artefacts that the ECA included in its narration with a high degree of accuracy. This finding, though, should be considered with caution as the analysis of the gaze trail data produced during the presentations showed that participants indicated that they saw objects, which did not cast gaze on the interface. A two-sided Fisher's exact test was performed to identify any associations between males and females in their ability to recognize objects on the interface. Furthermore, we wanted to see how the type of ECA (attention-grabbing vs. non-attention grabbing), the age of the participants, and the type of content, affected their ability to recognize objects.

Table 2. The results of the object recognition results

Order of presentation	Groups	AG (Yes/No)	NAG (Yes/No)
Simple/Complex	Females	25/20	38/7
	Males	35/10	24/21
Complex/Simple	Males	22/23	28/17
	Females	18/27	27/18
Total (Y/N)		100/80	117/63

Among multiple factors, we only found a significant association between the age of the participants and the object recognition responses (p=.025). No other significant associations were found.

3.2 Retention Performance

Table 3 shows the retention performance of the participants according to ECA type and order. The overall means indicate that participants performed better with the non-attention grabbing ECA. However, the effect of the manipulation of the various variables (e.g., age, gender, etc.) is not clear in the overall means. Hence, we conducted a series of 2 x 2 ANOVAs taking as independent variables a) the type of ECA, b) The participant's gender and age, c) the order of presentation and d) the type of content and as dependent variables a) the retention scores and b) confidence.

Table 3. Mean retention performances

	AG		NAG	
Order of presentation	Mean	SD	Mean	SD
AG(Simple)/NAG(Complex)	17.6	12.7	20.5	12.8
AG(Complex)/NAG(Simple)	13	13.9	23.8	11.9

The tests showed significant effects for both retention scores and confidence. With regards to retention scores, a correlation between the type of ECA and the participant's gender (F (1, 20) = 5.845; p < .05) (see Figure 2) was observed. This finding was further analyzed using simple main effects analysis.
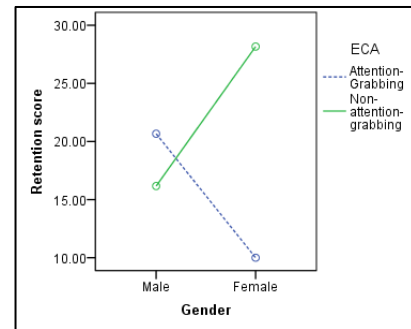

Figure 2. The interaction of retention scores for ECA and gender

It showed that the variation of ECA influenced the retention performance of the female participants (F (1, 20) = 7.509; p < .05) but not the retention performance of the male participants. A closer investigation of the descriptive statistics (see Table 4) reveals that the female participants scored significantly higher when they experienced the presentations with the non-attention-grabbing ECA (mean score = 28.1). The male participants had the exact opposite results. They scored better with the attention-grabbing ECA (mean score = 20.6).

Table 4. Retention performance

ECA	Gender	Mean	SD
Attention-Grabbing	Male	20.6	14.71
	Female	10.0	9.14
Non-attention grabbing	Male	16.1	8.68
	Female	28.1	12.31

The attention-grabbing strategies had no impact in attracting the female participants' attention back to the presentations. In order to find out which of the two strategies (or both) did not work, we conducted a 2 x 2 ANOVA, taking gender and attention-strategy as independent variables and retention scores as dependent. The test showed no significant effects of attention-strategy on gender. This shows that both attention-grabbing strategies (humorous and serious) had no impact on the female participants, which could explain their low retention scores. With regards to the possible effects on confidence, the following observations were made:

- The type of ECA (F (1, 20) = 17.440; p < .01)
- The order of presentations (F (1, 20) = 11.480; p < .01)
- An interaction between order of presentations, and type of ECA (F (1, 20) = 11.267; p < .01)
- An interaction between the type of content and the order of presentation (F (1, 20) = 17.440; p < .001) and finally,

- An interaction between the type of ECA and the type of content (F (1, 20) = 17.440; p < .01).

All participants rated the confidence of their answers in the retention tests lower with the attention grabbing ECA (mean confidence = 3.8). This shows that participants were actually unsure about whether the answers they provided were the right or not. The low confidence of the answers for the presentations with the attention-grabbing ECA can explain the low retention scores of the female participants. However, it does not explain the high-retention scores of the male participants. The confidence of their answers with the attention grabbing ECA is too low (mean confidence = 4.83) for the results they achieved. A further simple main effects analysis helped to determine the following:

- The variation of order influenced the participants who used the non-attention grabbing ECA (F (1, 20) = 22.747; p < .001) but not the attention grabbing ECA.

- The variation of order influenced the participants who experienced the complex content (F (1, 20) = 28.610; p < .001) but not the simple content.

- The variation of content influenced the participants who used the non-attention grabbing ECA (F (1, 20) = 22.747; p < .001) but not the attention grabbing ECA.

The significant effects above can be attributed to the group of the 35+ participants. The participants of that group scored the confidence in their answers to be very low (mean confidence = 3.66), the lowest from the participants of all other groups. This way, the overall confidence score of all groups in the second order dropped significantly creating the significant effects mentioned above.

3.3 Eye tracking – Heat maps

The eye-tracker produced individual heat maps (see Figure 3a) showing where the participants looked at and for how long. The hotter the area, the more it was noticed and looked at by the participants. The heat maps indicated that the background images attracted more attention than the ECA itself. This shows that the ECA was effective in directing the participants' attention to the objects/artefacts of the background images. However, the participants paid more attention to the background images, when they experienced the presentations with the attention-grabbing ECA than with the non-attention-grabbing ECA. An example of heat maps of a participant with similar effects is shown in Figure 3a. This effect is independent of the type of content (simple vs. complex) and the attention-strategy used. The most plausible explanation is that the attention-grabbing ECA attracted too much attention to it, and hence, distracted participants from paying attention to the content presented about the location. The constant request of the ECA for attention (regardless of whether the user was paying attention or not), most likely amplified the impact of this effect. The differences in attentiveness between the male and female participants were not significant. However, it is evident that in all four presentations, the elderly female participants paid too much attention to the presentations with the attention-grabbing ECA, more than the other participants. This attention could be interpreted as a sign of increased cognitive activity to follow the presentations, which could explain the negative retention performance of the older female participants.
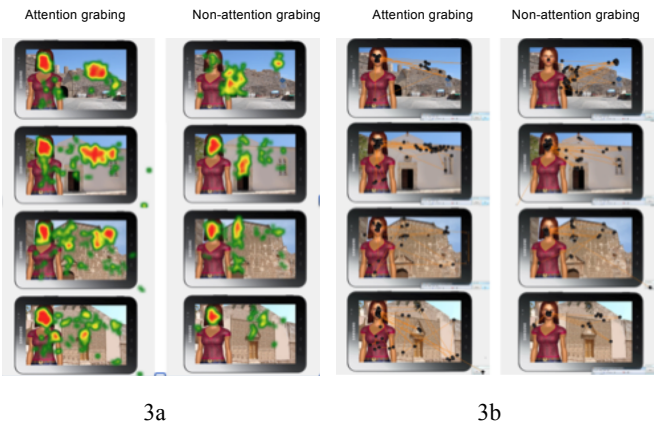

Attention grabing Non-attention grabing Attention grabing Non-attention grabing

3a 3b

Figure 3. Heat maps and Gaze trails of participants using both ECA systems

3.4 Eye tracking – Gaze trails

A gaze trail (see Figure 3b) shows the order in which participants looked each section of the interface and for how long. To aid with analysis the interface was divided into two look-zones, the background (including the floating window), and the ECA. So far, the findings indicated that the attention-grabbing ECA attracted more attention to itself than the non-attention grabbing ECA. This effect is independent of the attention strategy used (humorous vs. serious), and most likely distracted participants from paying attention to the background images. Participants using the non-attention grabbing ECA, most likely did not pay so much attention to the ECA itself. To verify these qualitative findings, a series of 2 x 2 ANOVAs, taking the number of fixations and fixation duration per participant as the dependent variables, and type of ECA and look zone as the independent variables, were conducted. A highly significant effect of ECA on the mean fixation duration (F (1,188) = 17.661; p < .001) and number of fixations (F (1,188) = 33.549; p < .001), and a highly significant effect of the look zone on the mean fixation duration (F (1,188) = 159.840; p < .001) and number of fixations (F (1,188) = 79.994; p < .001) was found.

Table 5. Overall fixation data

	ECA	Zone	Mean
Mean Fixation Duration	Attention-Grabbing	Background	54.8837
		Avatar	25.4947
		Total	40.1892
	Non-attention-grabbing	Background	43.3308
		Avatar	19.2862
		Total	31.3085
	Total	Background	49.1073
		Avatar	22.3904
		Total	35.7488

	ECA	Zone	Mean
Number of Fixations	Attention-Grabbing	Background	134.25
		Avatar	70.6042
		Total	102.4271
	Non-attention-grabbing	Background	89.5417
		Avatar	45.6458
		Total	67.5938
	Total	Background	111.8958
		Avatar	58.125
		Total	85.0104

It can be clearly seen from Table 5 that participants paid more attention to the system with the attention-grabbing ECA (mean number of fixations = 102.42, and mean fixation duration = 40.18 sec). Furthermore, all participants, regardless of the type of ECA they used, paid more attention to the background images, than the ECA itself. However, the attention-grabbing ECA was more effective in directing the participants' attention to the background (mean number of fixations = 134.2 and mean fixation duration = 54.8). Hence, the assumptions based on the gaze trail images that the attention-grabbing ECA distracted participants from paying attention to the background images was not supported. The above findings, contradict the attention-grabbing argument against the use of ECA on computer interfaces (e.g. [14]). An ECA that uses verbal and non-verbal means to communicate information can divert the user's attention away from itself and towards objects of interest in the interface. However, there is still the issue of how effectively this focus guidance was done and whether participants look at the objects/artefacts pointed by the ECA in the background. The matches between the objects participants said they saw in the questionnaires and those they fixated on the interface were examined. It was evident that there was a match between the questionnaires and the relevant gaze fixations for most of the objects/artefacts. For example, in Figure 5, the participant has cast her gaze upon the "Christian Cross" at the top of the church several times, which shows that she understood what she has seen and explains why she confirmed the object in the questionnaire. However, it is unknown whether the participant successfully connected the object/artefact with the information presented by the ECA. The inclusion of visual questions (e.g., what is the name of this object) in the retention tests could help us provide further insights into this question. Also, there is a mismatch between the questionnaires and the relevant gaze fixations. Some of the participants said they saw an object in the questionnaire, but they did not cast their gaze on the relevant objects/artefacts on the interface. A likely explanation about this mismatch is that participants confused these objects/artefacts with other similar objects/artefacts they saw (and cast their gaze upon) on the interface.

Figure 5. Fixations on an object confirmed in the questionnaire

3.5 Correlation (facial expressions, gaze trails, retention tests)

We performed a correlation between the facial expressions, gaze trails (number and mean fixation duration per section of the interface) and retention tests to help explain particularly good or bad performances in the retention tests. From both groups of participants, we chose one example of particularly good and bad performance in the retention tests. Then, based on the correlated data, we attempt to explain the outcome. Table 6, shows the retention performances of two participants, and the conditions under which they tested the systems. The retention samples selected for the attention grabbing conditions reflect both attention-grabbing strategies (humorous and serious). Beginning with the bad performances, it can be clearly seen that the selected participants remembered very little when they watched the

presentations with the attention-grabbing ECA (humorous or serious).

Table 6. Sample retention performances

Order of presentation	AG Mean	AG SD	NAG Mean	NAG SD
AG(Simple)/NAG(Complex)	17.6	12.7	20.5	12.8
AG(Complex)/NAG(Simple)	13	13.9	23.8	11.9

A careful examination of table 7, reveals that the attention-grabbing ECA attracted more attention to itself and the background than the non-attention grabbing ECA (measured in terms of total number and mean fixation duration). The difference between the two types of ECA was statistically significant only for the mean fixation duration ($F_{(1, 28)}$ = 5.436; $p < .05$). Furthermore, at each presentation until the interruption message occurred the participants had either neutral/blank or attentive facial expressions, which shows that they were attentive to the presentations. However, when the ECA requested for attention, the following observations were made:

a) The first time the ECA requested attention; participants were either surprised or curious, possibly because they did not expect the ECA to observe their behavior.
b) The initial emotion degraded gradually in every presentation when the ECA asked for attention. In fact, one of the two participants got annoyed at the second presentation when the ECA asked for their attention again. This was most likely because she was already paying attention to the presentation.

Table 7. Mean fixations and times of participants with bad performances

ECA	Background Number of Fixations Mean	SD	Background Mean Fixation Duration Mean	SD
AG	125.2	48.6	59.4	14.4
NAG	102.2	57.2	43.1	25

ECA	ECA Number of Fixations Mean	SD	ECA Mean Fixation Duration Mean	SD
AG	66.8	29.7	29.9	8.9
NAG	52.5	36.6	19	13.4

With regards to good performances, participants remembered a moderate amount of information from the presentations (see Table 4). A careful examination of the data in the table reveals the following about the two participants. The interaction between the type of ECA and the look zone (i.e., background and avatar) was significant for the total number of fixations ($F_{(1, 28)}$ = 8.007; $p < .01$). A simple main effect analysis showed that the attention-grabbing ECA attracted significantly more attention to the background ($F_{(1, 28)}$ = 5.939; $p < .05$) than the non-attention grabbing ECA. The mean number of fixations differences for the type of ECAs did not reach statistical significance levels. Until the interruption message of the attention-grabbing ECA, participants had a neutral/blank face, which shows their attentiveness to the content of the presentations. Nonetheless, in contrast to the other two sets of participants, their facial expressions during the interruption messages were more constrained (e.g., a slight smile, or even neutral/blank).

Table 8: Mean fixations and times of participants with good performances

ECA	Background			
	Number of Fixations		Mean Fixation Duration	
	Mean	SD	Mean	SD
AG	132.2	43.1	51.4	14.07
NAG	89.3	43.9	42.3	20.0
ECA	ECA			
	Number of Fixations		Mean Fixation Duration	
	Mean	SD	Mean	SD
AG	28.4	10.8	21	14.1
NAG	56	32.3	23	10.0

This most likely means that those participants were not distracted by the attention-grabbing messages and they were able to keep their focus on the presentations. In addition, the patterns observed in the participants with the bad performances when the ECA requests for attention (see above) were not observed in these participants.

4. CONCLUSION AND FUTURE WORK

This study revealed that an ECA should not attract attention to itself more than necessary, to avoid becoming a distraction from the flow of the content. We found proof that an ECA with attention-grabbing capabilities (humorous or serious), can effectively divert the participants' attention focus to relevant content-objects in the background. We also found that the attention-grabbing messages had a detrimental effect on the overall retention performance. This affected the female participants more than the male participants. The male participants performed better with the attention-grabbing ECA, while the female participants did so with the non-attention grabbing ECA. Therefore, even when the system can actually react to the user's attention state, a minimum threshold where the ECA can request for the participant's attention must be established. This threshold is difficult to determine as the ECA must avoid becoming tiresome, but also must be effective enough to attract the participants' attention back to the presentation when it has deviated. The multiple communication channels used by the ECA result into an increased cognitive overload that makes it difficult to readily follow the content of the presentations. We argue that for those participants who are overloaded it is much easier to use more "traditional" channels of communication when designing mobile guide systems, such as text and voice. Overall, this study produced viable evidence to support our claim that our proposed method for evaluating the accessibility of cultural heritage content actually works. However, given the limitations of the study (subject and expert analyst pool size) this claim should be considered with caution, as it cannot be considered generally applicable to all ECA-based mobile information systems for cultural tourism. Additional research is needed with a bigger subject pool size where the participants would experience the ECA's through an actual mobile device (tablet or smartphone). Furthermore, additional experts are needed to ensure better quality of video coding and validation of facial expressions. Using SHORE to attach an intensity range (from 0 – 100 with a higher value meaning more intense expression) to facial expressions could certainly strengthen the validation process. Finally, quantitative analyses of the facial expressions could provide further insights about the user's emotional state as s/he experienced the different types of ECAs and types of cultural information.

5. REFERENCES

[1] Witkowski, M., Arafa, Y., and de Bruijn, O. 2001. Evaluating User Reaction to Character Agent Mediated Displays using Eye-tracking Equipment, *In: Proc. AISB'01, Symp. on Information Agents for Electronic Commerce*, 21 – 24 March 2001, University of York, UK, pp. 79-87.

[2] Takeuchi, A., and Naito, T. 1995. Situated Facial Displays: Towards Social Interaction. *Human Factors in Computing Systems: CHI'95 Conference Proceedings*, May 7 - 11, 1995, Denver, Colorado, USA, pp. 450-455.

[3] Eichner, T., Prendinger, H., André, E. and Ishizuka, M. 2007. Attentive presentation agents. *Proc 7th Int'l Conf on Int'l Virtual Agents (IVA'07)*, Springer LNCS 4722, Paris, France, Sept. 2007, pp. 283-295.

[4] eMarketView 2013. Eye tracking services (Available from: http://www. emarketview.com/ servicios/usabilidad-conversion/eye-tracking) [Accessed February 19 2013].

[5] Joumana, M., 2011. Emotions and Face expressions. [online image] Available from: http://www.cedarseed.com/ [Accessed July 15 2011].

[6] Fraunhofer Institute for Integrated Circuits IIS, **"**Sophisticated High-speed Object Recognition Engine (SHORE)**",** version 6.30, computer program and manual, Erlangen, Germany.

[7] Grafsgaard, J.F., Boyer, K.E. and Lester, J.C., 2011. Predicting Facial Indicators of Confusion with Hidden Markov Models.;In Proceedings of J. Spatial Information Science. 2011, 97-106.

[8] Simple Usability Inc., 2013. Emotion Response Analysis through EEG technique (Available from: http://www.simpleusability.com/) [Accessed February 19 2013].

[9] Microsoft Visual Studio 2010 for Microsoft Windows 7, software, version 2010 RTM (10.0.30319.1), Microsoft Corporation (available from http://msdn.microsoft.com/en-us/vstudio/default.aspx)

[10] Haptek Player for Microsoft Windows 7 software, version 4, Haptek Inc. (available from http://www.haptek.com/)

[11] EyeTellect's gaze tracking software for Microsoft Windows 7, software version unknown (available from http://www.eyetellect.com/gazetracker)

[12] Sauro, J. (2014). Nine Misconceptions about Statistics And Usability. [Blog] *Measuring Usability*. Available at: http://www.measuringusability.com/blog/stats-usability-errors.php [Accessed 1 Sep. 2014].

[13] Sauro, J. and Lewis, J. (2012). *Quantifying the user experience*. 1st ed. Amsterdam: Elsevier/Morgan Kaufmann.

[14] Walker, J. H., Sproull, L., & Subramani, R. (1994). Using a human face in an interface. *In Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*, pp. 85-91.

[15] Trac.media.mit.edu, (2014). *MindReader API*. [online] Available at: http://trac.media.mit.edu/mindreader/ [Accessed 2 Sep. 2014]