A Mobile Interactive Robot for Gathering Structured Social Video

Alexander Reben
MIT Media Lab
75 Amherst St
Cambridge, MA 02139
areben@media.mit.edu

Joseph Paradiso
MIT Media Lab
75 Amherst St
Cambridge, MA 02139
joep@media.mit.edu

ABSTRACT

Documentaries are typically captured in a very structured way, using teams to film and interview people. We developed an autonomous method for capturing structured cinéma vérité style documentaries through an interactive robotic camera, which was used as a mobile physical agent to facilitate interaction and story gathering within a ubiquitous media framework. We sent this robot out to autonomously gather human narrative about its environment. The robot had a specific story capture goal and leveraged humans to attain that goal. The robot collected a 1st person view of stories unfolding in real life, and as it engaged with its subjects via a preset dialog, these media clips were intrinsically structured. We evaluated this agent by way of determining "complete" vs. "incomplete" interactions. "Complete" interactions were those that generated viable and interesting videos, which could be edited together into a larger narrative. It was found that 30% of the interactions captured were "complete" interactions. Our results suggested that changes in the system would only produce incrementally more "complete" interactions, as external factors like natural bias or busyness of the user come into play. The types of users who encountered the robot were fairly polar; either they wanted to interact or did not - very few partial interactions went on for more than 1 minute. Users who partially interacted with the robot were found to treat it rougher than those who completed the full interaction. It was also determined that this type of limited-interaction system is best suited for short-term encounters. At the end of the study, a short cinéma vérité documentary showcasing the people and activity in our building was easily produced from the structured videos that were captured, indicating the utility of this approach.

Categories and Subject Descriptors

H.5.2 User Interfaces (D.2.2, H.1.2, I.3.6)

General Terms

Design, Experimentation, Human Factors.

Keywords

HRI, Social Robotics, Interaction, Automatic Documentary

1. INTRODUCTION

A robotic camera called Boxie was built that had a goal of actively capturing a story about its environment and the people within it. That is, that the robot had a specific story capture goal and leveraged its mobility and interactivity together with the capabilities of the ubiquitous sensor network to achieve that goal. This device also allowed for an active first person view that many passive distributed media capture systems don't enable [1]. It can also reach areas that 3rd person systems do not cover by its inherent mobility. In these blind areas the robot either recorded data there or prompted others to move it to an area where sensors are active.

The novel approach to this system is that it enables activities of interest to be effectively and actively captured by leveraging human empathy and interaction as a social robot. Sociable robots can be defined as robots which leverage social interactions and cues in order to attain the internal goals of the robot [2]. Through this empathy and interaction, this engagement encouraged the person interacting with the robot to share their story, in a meaningful way. Similar robots have been developed that leverage humans to achieve simple goals using empathy [3]. We evaluated the effectiveness of different forms of interaction which were developed on the robot platform. This interaction design and testing in real world scenarios was the focus of the investigation. The robot acts as a facilitator to coax those who may not initially be inclined to interact with the system to share their stories. The documentary that was created started small scale, at the immediate level of the person interacting with the robot, and expanded through that interaction to encompass stories from others who subsequently encountered the robot. As the style of the documentary is cinéma vérité, it provokes the subjects.

A narrative "thread" followed the robot through its interactions and path of story goal achievement. Through this, we can see how the robot's users are connected within the story line, through the robot's interactions. This provides a way to incorporate social interaction within the framework of ubiquitous media systems. This system is the first we know of that has the specific goal of actively capturing a documentary within a ubiquitous media environment by leveraging human intelligence, interaction and emotion. This novel approach created a rich story capture system within and outside of ubiquitous media frameworks.

2. INTERACTION DESIGN

As with all social robots, interaction design plays a pivotal role in their success. The first step in the design of a sociable robot that

leverages humans is to consider the interaction with the user. There are several factors that determine the success of such a system. The factors for our particular system were overall "cuteness", interaction simplicity, appearance and behavior.

2.1 Appearance

Appearance is an important factor in addressing interactivity and usability of a robot. Since the robot would be relying on its cuteness in order to leverage humans, special care was taken in the design of its appearance. The overall aesthetic of the robot was box-shaped with a bottom mounted track drive. We chose this shape because of its simple lines and familiar appearance. The physical characteristics of most importance included eye size, eye spacing and position, head size, and body proportions. The eyes of the robot were chosen to be large and circular. Geldart et al. determined that figures with larger eye sizes were found to be more attractive by both adults and children [4]. This eye geometry yielded the "cutest" looking results in the prototypes. The eyes were set far apart and low on the head, making the robot appear naive and young. The head was made large in relation to the torso, a 3:2 ratio was found to be most aesthetically pleasing. A short squat body was the most childlike configuration (see Figure 1). We also considered how the user would hold the robot. The only way to test how a user might approach holding the robot was to develop physical prototypes and ask users to grasp the robot in the most natural way.

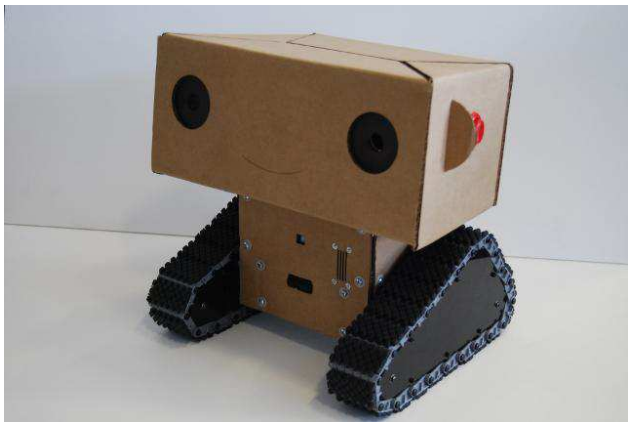

Figure 1. Boxie the robot.

2.2 Subject Acquisition

The overall interaction paradigm for the robot was that of "active" story gathering. Bergström et al. determined that an active robot in a crowded public area (that of a shopping mall) was most successful in eliciting interactions with passers-by if it took an active role in the human acquisition process [5]. It was also found that humans will adjust their positions in order to be "noticed" by the face of the robot. The idea of a human being "noticed" was a central consideration in the interaction design of the robot. To be noticed, the robot needed a face and a gaze. It also needed to recognize the user and acknowledge their presence. This also placed the camera embedded in the robot's eye in a position to have a face-to-face interview with the subject.

2.3 Engagement

The main mechanism the robot uses to leverage humans is cuteness. Our emotional reactions and empathy for cute things is hardwired into our brains through evolution. The subject of cuteness has been long studied by evolutionary psychologists. The reaction we have to cute things originates with the need to care for babies [6]. The cuteness of babies motivates parents to care for and protect them. We used the cuteness of the Boxie robot to draw users in and keep them engaged. The cuteness paradigm we chose to use is that of the lost and helpless child. It was felt that if the robot was perceived as a helpless and lost "child-bot" that user's instincts would engage and they would interact with the robot until they were able to help it. A major consideration was how to engage the user and keep them engaged without making the user feel suspicious of the robot.

"…the rapidity and promiscuity of the cute response makes the impulse suspect, readily overridden by the angry sense that one is being exploited or deceived." [7]

We avoided this pitfall by carefully molding the interaction with the robot to gradually ask the user to work more for the system. The robot also interacted in a two-way fashion, giving the user something in-kind to their input. For example, when the user told the robot something about themselves, the robot would tell it something about itself. With this approach, we avoided the perception by the user that they were being used by the system, making the interaction enjoyable and effective.

2.4 Behavior

Creating behavior that was "cute" was not as immediately apparent as designing the physical appearance. The behavior of the robot needed to be scripted, just as an actor would be. The particular behavior of the robot depended on its motive and its current situation. Depending on the goal of the system, the robot modified its behavior to best achieve the objective. The voice and movement of the robot were the main avenues through which to represent behaviors.

Through movement, we could play with the personality of the robot dynamically. Non-verbal behavior can express as much as and augment the content of verbal communication [8]. Movement is an important factor for displaying expression in robots with constrained appearances such as Boxie [9]. The nominal movement behavior of the robot was to seek out people while getting stuck or lost. The robot was made to act naive by appearing to not know where it was going. By appearing lost, people around the robot would be compelled to help it, just as they would with a lost child or helpless animal. When the robot got itself suck, its behavior was reminiscent of a wounded animal or a child in need. For example, if the robot wedged itself in a corner or under an obstacle, it squirmed around back and forth as if it was trapped. The robot would detect if it was stuck by using its distance sensor and accelerometer. When the robot got itself stuck, it would raise slightly off the ground, which could be detected by an increase in acceleration due to gravity and an increase in the robot to floor distance. It would wait to detect

Aera Chair: Aisling Kellihera
ACM Multimedia'11, Nov 28 - Dec 1, Scottsdale, Arizona, USA.

reflected body heat of a nearby human or time-out in software and move again. When the robot came to a complicated intersection, it moved back and forth as if it was searching or confused about what it should do. This confusion led people to believe the robot was lost or helpless to find its way. The robot slowed down before this behavior to mimic hesitance or fear of the path ahead. When the robot sensed a human it stopped in the direction of the human and motioned toward them, indicating that it would like to initiate a conversation. This indication of conversational intent was an important factor in the capture of a human for interaction [10]. Past research has found that behavior plays a crucial role in attempting to get a user to perform an action for a system. One robot that did produce a tangible output from its users was shaped like a trash can, and would rove around trying to elicit children to deposit trash. The robot could not complete its goal on its own without the interaction of the children. The robot utilized vocalizations and behaviors to try to elicit trash from the children. It was found that the robot could achieve its goal if the robot was moved toward an area with trash by the children [11].

2.5 Story Capture

The type of story we chose to capture was the documentary. We crafted a script for the robot to speak in order to capture a story about the current place the robot was in. The childlike voice of the robot was scripted in such a way as to complete the robot's goals. Two scripts were produced and implemented. One involved giving the user simple commands while the other added personality and provided a two way conversation. The intent was to study the difference in user responses in comparing the types of scripts implemented. In "Relational agents: a model and implementation of building user trust" it was found that adding "small talk" to the interaction paradigm of an embodied agent increased the user trust for that agent [12]. At the beginning and end of each script, the robot would make an edit point in the internal camera. This separated the individual interactions inside the robot, effectively pre-editing and sorting the footage. The scripts interviewed those that the robot found so that the video shot would be able to be assembled into a coherent documentary. The place in the robot's script was advanced by the subject acknowledging they were done speaking by pressing a button on the side of the robot's head. The interviewees would speak directly to the robot looking at the robots "face" (hence camera) during the capture process (Figure 2). (see reference [13] for detailed scripts).

Figure 2. Example of face-to-face interaction from the view of the robot.

When the robot was finished roving for the day, it was retrieved and the video was downloaded along with the data to indicate how to edit to documentary together. This data was used to form a video showing the robot's journey and its interaction with people.

3. RESULTS

3.1 Interaction Time

The minimal amount of time to complete the full useable interaction sequence was a little over 3 minutes. A full useable interaction sequence was defined as interaction with the robot to the point of being a useable narrative (over 90% question completion rate). In a sample group of 15 subjects, a clear separation between full interaction and partial interaction was seen at the 2.5 to 3 minute mark [13]. We used this mark to identify the videos which were full useable interaction vs. partial interaction. There exists a large spike of interactions which lasted for less than 1 minute. The majority of these interactions were found to be, through analysis of the video and question progression, either users pushing the button and walking away or not choosing to continue interacting with the robot after quickly investigating it. It was found the 70% of the interactions fell into this "partial interaction" category while 30% fell into the "full useable interaction category" [13]. Note that it was to be expected that the number of complete interactions would be low. This was due to the fact the complete interactions took a disproportionally large amount of time vs. incomplete interactions. Since the time the robot was roaming was finite, more incomplete interactions could be recorded vs. the number of complete ones. Furthermore, complete interactions implied a full commitment to the robot, which required more time of the participant.

3.2 Robot Abuse

An interesting trend emerged; users who did not complete the full interaction also tended to mistreat the robot – i.e. handle it roughly. We measured how mistreated the robot was by using simple accelerometer data to infer how it was handled, hit or shaken. While users who reached the full interaction stage attained an average maximum G force of 2.4 (Figure 3), those who did not complete the interaction fully reached an average maximum of 4.9 Gs, over twice as much (Figure 4)[13]. While 2.4 Gs is within the limit of normal acceleration of the robot (associated with handling and tilting), a force of 4.9 Gs indicates rough handling and a few interactions where the robot was placed back on the ground roughly. There is a clear split between the interaction types. This data could be used to sense a bad interaction for either automatic editing or for the robot to take action and try to evade this type of person.

3.3 Documentary Production

The video gathered from the robot was able to be used to generate a coherent documentary about the spaces it had visited. After a week of roaming a space, we were easily able to extract enough footage to generate an edited documentary lasting 5 minutes 25 seconds [13]. We captured a total of 50 clips which equated to roughly an hour of raw footage, about a third of which were useable. We were able to use the footage from both partial and full interactions, with each type lending their own character to the story. Because of the way in which the raw video was automatically sorted by the button pushes of the users after each segment of interaction, this video [14] was straightforward to assemble and has been well received by all who have viewed it.

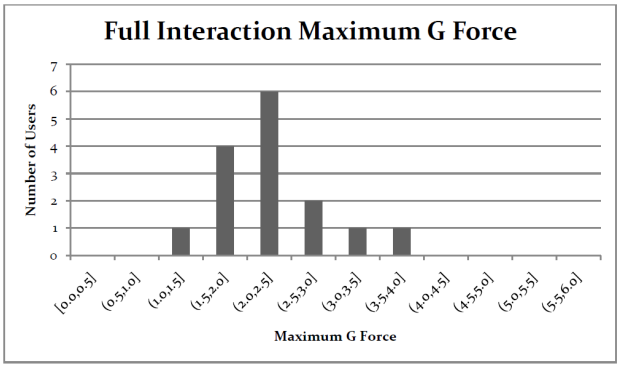

Figure 3. Distribution of maximum G forces encountered during full interactions.

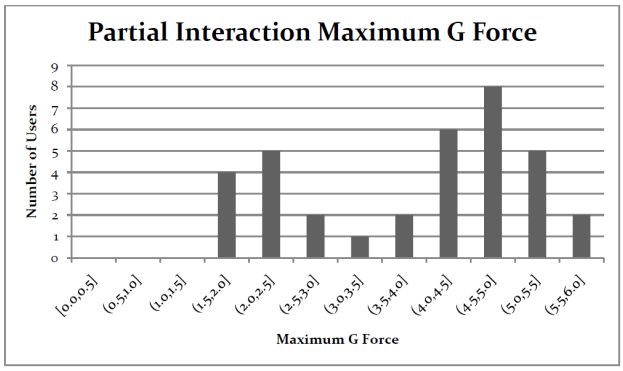

Figure 4. Distribution of maximum G forces encountered during partial interactions.

4. CONCLUSIONS

The creation of a physical agent to facilitate interaction resulted in a successful story-capturing robot, which effectively engaged people to extract structured interviews about its environment. Success was measured as the ability to create a coherent cinéma vérité style documentary with the robot's footage. It was found that 30% of the interactions captured were "complete" interactions. Results suggested that changes to the system would only produce incrementally more "complete" interactions, as external factors like natural bias or busyness of the user come into play (as it was set loose in an active workplace). The types of users who encountered the robot were fairly polar; either they wanted to interact or did not - very few partial interactions went on for more than 1 minute. It was also determined that this type of limited-interaction system is best suited for short-term encounters. At the end of the study, a coherent movie was easily produced from the video clips captured, proving that their content and organization were viable for story-making.

Some suggestions can be made for the development of future agent-based physical story capture systems. Keep the interaction as interesting as possible as users are more likely to share stories if the agent also shares back with the user. Anthropomorphize the system to create a connection with the person you are trying to leverage. Have the agent offer them something in exchange for their interaction. Make the agent seem like it needs the user. Be transparent with the purpose of the agent, as users are prone to be skeptical about a system that needs them but does not tell them why. Try not to be annoying since there is a fine line between attempting to capture a user and annoying them; the system is most effective just before that line. Look "good" because the agent should look the part. This is as important a consideration as the technology inside.

5. ACKNOWLEDGMENTS

We acknowledge the MIT Media Lab and its research sponsors for supporting this work, together with our colleagues, particularly Mat Laibowitz, Nan-Wei Gong and the rest of Responsive Environments and Sheng-Ying Pao for their help and support.

6. REFERENCES

[1] M. Laibowitz, 2010. *Distributed narrative extraction using wearable and imaging sensor networks*, PhD. Thesis, MIT Media Lab.

[2] C. Breazeal, 2004. *Social interactions in HRI: the robot view*, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 34, pp. 181-186.

[3] K. Kinzer, Tweenbots, Accessed 2011, http://www.tweenbots.com/

[4] S. Geldart, et al., 1999. "Effects of eye size on adults' aesthetic ratings of faces and 5-month-olds' looking times," *PERCEPTION-LONDON-*, vol. 28, pp. 361-374.

[5] N. Bergström, et al., 2007. Modeling of natural human-robot encounters," School of Computer Science and Engineering, Royal Institute of Technology, Stockholm, Sweden, 2008. Tavel, P. *Modeling and Simulation Design*. AK Peters Ltd., Natick, MA.

[6] D. Dutton, 2003. "Aesthetics and evolutionary psychology," *The Oxford Handbook of Aesthetics*, pp. 693–705.

[7] N. Angier, "The Cute Factor," in *The New York Times*, ed. New York, 2006.

[8] B. DePaulo, 1992. "Nonverbal behavior and self-presentation," *Psychological Bulletin*, vol. 111, pp. 203-243.

[9] C. L. Bethel and R. R. Murphy, 2006. "Affective expression in appearance constrained robots," presented at the *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, Salt Lake City, Utah, USA.

[10] S. Satake, et al., 2009. "How to approach humans?: strategies for social robots to initiate interaction," presented at the *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, La Jolla, California, USA.

[11] Y. Yamaji, et al., 2010. *STB: human-dependent sociable trash box*, presented at the Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction, Osaka, Japan.

[12] T. Bickmore and J. Cassell, 2001. "Relational agents: a model and implementation of building user trust," presented at the Proceedings of the SIGCHI conference on *Human factors in computing systems*, Seattle, Washington, USA

[13] A. Reben, 2010, *Interactive Physical Agents for Story Gathering*. Master's thesis. MIT Media Lab, Cambridge, MA, USA.

[14] A. Reben, Boxie, Accessed 2011, http://boxie.media.mit.edu