Designing the Human-Robot Interaction for PiNAOqio

Periklis Korkontzelos (pk1913), Imperial College London

Abstract

This report consists of an overview of the Human-Robot Interaction (HRI) between the user and PiNAOqio, an educational robotic companion with the ability to read hard-copy books. The expected and desired outcomes from the HRI include the improvement of information retention and user experience, as an effect of increased interaction. PiNAOqio is designed to perform gestures that enrich the learning process and to emulate human storytellers in order to provide an activity to which the user is accustomed. The implementation will be carried out using the NAOqi API in the Python SDK. The different stages have been designed to ensure the validity and consistency of the results so that meaningful conclusions can be drawn.

I. Introduction

The aim of this project is to evaluate the user experience and information retention of a person being read a story, while varying the level of interaction between the user and the storyteller. The possible methods of interaction consist of (a) gestures made by the PiNAOqio to facilitate meaning while telling the story and (b) an interrupt routine made by the user through an external microphone to ask for clarifications or further details. It is hypothesized that by varying the level of interaction, information retention and user experience change accordingly (Laguerta, Korkontzelos, Warner & Rusu, 2016). Thus, understanding how different designs accompany different interactions and how these interactions affect the emotional and psychological state of the user is crucial in evaluating and comparing the different models.

The aim of this report is to discuss the implementation of the HRI and how this will differ from a non-interactive solution by enhancing user experience and information retention. This will be done by outlining the system requirements of the HRI subsection of PiNAOqio, discussing the reasoning behind the choices made and finally, covering the implementation and technical details.

II. System Requirements

In order to build a robust system, careful consideration should be put into the design of the HRI. The interaction should be consistent for all case studies and users to ensure that meaningful correlations can be drawn from the results gathered. To simplify the implementation, HRI can be divided in two subsections, namely, PiNAOqio-user interaction (PUI), whereby PiNAOqio tries to reach out to the user through speech or action and user-PiNAOqio interaction (UPI) whereby the user can interrupt the storytelling and talk to the robot through an external microphone. This distinction simplifies the process of designing the HRI as different requirements can be set for each subsection, addressing specific difficulties and targeting different goals set for each stakeholder.

a) PUI

From the point of view of the robot there are two main objectives that if satisfied, would render the PUI successful. The primary objective is to emulate a human storyteller as closely as possible in order to increase the comfortability perceived by the user. Per the *Social Agency Theory,* the combined effect of multiple interaction techniques exceeds the individual effect of each one (Ham, Bokhorst & Cabibihan, 2011). Therefore, a set of different parameters such as gazing pattern and body posture are put together to create a strong, positive impression to the user. At the same time, past research has investigated the effect of different gazing profiles on information retention of participants (Mutli, Forlizzi & Hodgins, 2006). The results suggested a positive correlation between the amount of the story remembered and length of time being gazed, which renders gazing a desirable way of interacting.

Further, it is important that the gestures and body language of PiNAOqio accompany his speech in order to facilitate meaning. An earlier study investigated the positivity in the message of a speaker as interpreted by the listener and was found that gestures linked to speech incur the most benefit, compared to random or no gestures at all (Maricchiolo, Gnisci, Bonaiuto & Ficca, 2009). The same result is supported from a physiological research in which participants were exposed to speech-associated gestures and their brain activity was analysed using network analysis of neuroimaging data. Results suggest that gestures associated with speech carry semantic information that is being used by the listener to clarify ambiguities arising with spoken discourse (Skipper, Goldin, Nusbaum & Small, 2007). Therefore, it is important not only to make the robot move while it is narrating, but to have specific actions that accompany key words or phrases to create meaning in the context of the story being read.

Earlier research suggests that pointing gestures implemented by a robot providing spatial information can significantly improve speech comprehension (Cabibihan, So & Nazar, 2009). In the context of a story therefore, gestures can be greatly beneficial in forming a highly detailed representation of the setting.

However, implementing an interactive robot does not only serve to increase the amount of information passed to the user but also improve his experience. A research studied the effectiveness of using a robot companion to help in the navigation of wheelchair users (Sarabia & Demiris, 2013) and found that users much preferred the robot over the simulator (30% difference when it came to the favourite driving aid). A different study investigated the feeling of co-presence and involvement in a dyadic conversation and assessed it comparatively between a video conference, a digital avatar and an audio conversation (Garau, Slater, Bee & Sasse, 2001). On both co-presence and involvement scales the video conference was rated the highest, suggesting the participants' will to observe rather than just listen to their conversation partner.

b) UPI

In addition to designing a robust PUI system it is equivalently significant to develop a UPI system that allows the user to feel in control of the storytelling experience. The functionality of the external microphone was added to make the storytelling less stressful. Emotionally stressful environments are counterproductive because they reduce student's ability to learn (Sylwester, 1994) and are more likely to contribute to an overall negative experience.

At the same time, it is expected that by allowing the user to interact with the robot the information retention will also increase. Under normal operation, the user is likely to become distracted and miss parts of the story. However, if he is in a position to ask the robot to repeat or explain the part that was missed, as he would in a human-human interaction, then the amount of information retained can be significantly boosted.

III. Implementation

The simplest solution for programming the HRI would be to use the NAOqi API which is largely documented and compatible with the NAO robot. Other APIs do exist, for instance the *MoveIt! API* through ROS (ROS, 2016) but offer no significant advantage. The Python SDK can be used for development and it is the standard way of the programming the NAO with python. It contains enough functionalities and reduces development time, thus being the development platform of choice.

Similarly to the requirements specifications, the implementation of the HRI for PiNAOqio can be subdivided into two parts, the implementation of the PUI and UPI. Allowing the user to interact with the robot is relatively straightforward, requiring solely an external microphone. This has a wired connection to the raspberry Pi, which does the processing and transmits data to PiNAOqio via an Ethernet connection. At the beginning of each testing session, the user is given a set of keywords that can be used to alter the reading profile. Currently, the robot is programmed to accept commands that change the reading profile to: speak faster/slower, repeat/skip a section defined by the user (line/page) and go to a specific section. Added functionalities, such as varying the pitch of the voice, may be added in the future.

When the microphone input is activated, PiNAOqio runs a speech recognition algorithm to determine what has been said by the user. Then, a matching algorithm tries to pair the words interpreted with the ones from the dictionary of words/phrases stored in memory. Once a link has been identified, PiNAOqio adjusts the reading profile accordingly and continues reading.

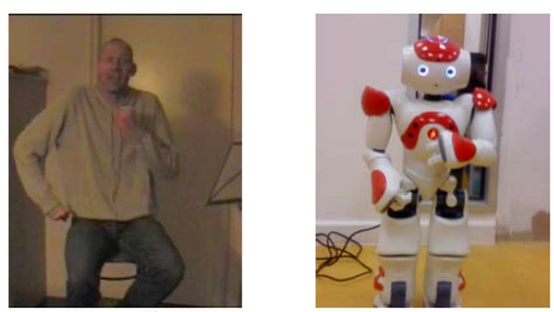
The PUI is slightly more complex as it consists of background motion that is running continuously and definite motion that is triggered upon the occurrence of a specific event. Most of the background motion is designed to match the style and mood of the story being read. The first parameter that can be manipulated is the posture of the robot, which can be used to reflect the emotional state of the narrator. For instance, the *Sitting* posture, in which the NAO sits with "buttock in contact with ground and torso upright" (Aldebaran 2.1 Documentation, 2016), can be used to indicate a stressed or fearful situation. Contrastively, the *Lying Belly* posture, in which the NAO lies "stretched facing down" (Aldebaran 2.1 Documentation, 2016) suggests a happier and more relaxed state.

In addition to selecting postures, there exists an API that allows for the programing of autonomous moves. These can be very useful in making the storytelling more realistic by filling the gaps that might occur during the experiment. In a study, it was found that people rely on body movements in addition to hearing to establish the speaker when a conversation is being initiated. (Vertegaal, Slagter, van der Veer and Nijholt, 2001). Thus, the Expressive Listening API is called whenever the user interrupts to talk, making PiNAOqio lean towards him to indicate that it is listening. A continuity in the bodily expressions of the robot increases the perception of the robot as a dependable, animate companion (Breazeal, 2009). The final background motion that is going to be set is the *Breath Configuration* API (Aldebaran 2.1 Documentation, 2016). Both the amplitude and rate of breathing can be programmed and varied to reflect changes in tone of narration and suggest the psychological state of the main characters, similarly to a human narration.

Custom motions accompany certain key words or phrases that are being read from the story. A dictionary is built prior to experiments and consists of several entries that can be clearly associated with an action. When an entry from the dictionary is read, the corresponding action is called. For example, figure 1 shows how possible gestures could be associated with 'shouting' and 'running' by a human and the NAO (Ham, J., Bokhorst, R. & Cabibihan, J. 2011).

Storyteller Nao

(a) "… and he shouted out; Wolf!"



Storyteller Nao

(b) "… as the villagers came running up the hill…"

Figure 1 - Examples of gestures by human and NAO

For the implementation of custom motions, the *ALMotion API* will be used which operates at 50Hz (Aldebaran 2.1 Documentation, 2016). This includes a range of libraries that can be used to program arbitrary movements. *Joint Control* allows the motion of each joint separately to create custom motion patterns. Further, *Stiffness Control* imposes a motor torque limitation that affects the speed and intensity of each action and can be used to affect the tone of the story.

IV. Conclusion

PiNAOqio is a robot companion for children that educates through storytelling while providing an enjoyable experience. As such, it should be designed with high levels of interaction allowing both the robot and the user to interact with their counterpart. The different modes of interaction are being described and the design choice are justified by referring to relevant literature. Finally, the implementation is outlined, explaining how the utilization of the NAoqi API contributes to a more sophisticated interaction.

V. References

1. Aldebaran 2.1 Documentation. (2016). *Naoqi Apis*. Available from: Doc.aldebaran.com
2. Bartneck, C., Kulic, D., & Croft, E. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence and Perceived Safety of Robots. *International Journal of Social Robotics, 1*(1), 71-81.
3. Breazeal, C. (2009). Role of expressive behaviour for robots that learn from people. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), pp.3527-3538.
4. Bromberg, N., Techatassansasoontorn, A. & Andrade, A. (2013). Engaging Students: Digital Storytelling in Information Systems Learning. *Pacific Asia Journal of the Association for Information Systems*, 5(1), pp.1-22.
5. Cabibihan, J.J., So, W.C., & Nazar, M.A. (2009). Pointing Gestures for a Robot Mediated Communication Interface. *International Conference on Intelligent Robotics and Applications*, Singapore.
6. Evans, M. & Rubin, K. (1979). Hand Gestures as a Communicative Mode in School-Aged Children. *The Journal of Genetic Psychology*, 135(2), pp.189-196.
7. Garau, M., Slater, M., Bee, S., & Sasse, M.A. (2001). The impact of eye gaze on communication using humanoid avatars. *The CHI 2001 Conference on Human Factors in Computing Systems*, Seattle, Washington.
8. Ham, J., Bokhorst, R. & Cabibihan, J. (2011). The influence of gazing and gestures of a storytelling robot on its persuasive power. *International conference on social robotics*.
9. Hostetter, A.B. (2011). When Do Gestures Communicate? A Meta-Analysis. *Psychological Bulletin, 137*(2), 297-315.
10. Kleinke, C.L. (1977). Compliance to requests made by gazing and touching experimenters in field settings. *Journal of Experimental Social Psychology, 13*(3), 218-223.
11. Laguerta, X., Korkontzelos, P., Warner, N., & Rusu, M. (2016). PiNAOqio: Story-telling for children using physical books and HRI. *Imperial College*, London.
12. Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior. *IEEE-RAS Conference on Humanoid Robots (Humanoids'06)*, Genova, Italy.
13. ROS, 2016, *MoveIt! Code API,* Available from: http://moveit.ros.org/code-api/
14. Sarabia, M., & Demiris, Y. (2013). A Humanoid Robot Companion for Wheelchair Users. *ICSR,* London.
15. Skipper, J., Goldin-Meadow, S., Nusbaum, H. & Small, S. (2007). Speech-associated gestures, Broca’s area, and the human mirror system. *Brain and Language*, 101(3), pp.260-277.
16. Sylwester, R., 1994. How emotions affect learning. *Educational Leadership*, *52*, pp.60-65.
17. Vertegaal, R., Slagter, R., van der Veer, G., & Nijholt, A. (2001). There is more to Conversational Agents than meets the Eyes. Paper presented at *The CHI 2001 Conference on Human Factors in Computing Systems*, Seattle, Washington.