PiNAOqio: Story-telling for children using physical books and HRI

Xavier Laguarta Soler (xl5512), Periklis Korkontzelos (pk1913), Nathan Warner (njw13), Mihnea Rusu (mr3313)
Imperial College London

Abstract: Recent advances in technology have led to a decrease in the reading of physical books in children [1]. Overuse of modern devices (e.g. smartphones, tablets) from an early age poses long-term health risks [2].

Our aim is to combine the old and the new; physical books with human-robot interaction (HRI) as a novel form of entertainment and education. We propose using a friendly miniature humanoid robot in conjunction with a page-turning mechanism to read books and interact with children. The system shall be evaluated through measurement of the level of information retention and user experience across different levels of HRI.

I. INTRODUCTION

In circumstances where a human story teller is not available the only alternative for small children would be an audio book. Due to the absence of interaction and human qualities audio books are not well received by children with limited attention spans. PiNAOqio offers to resolve this issue by actively engaging and interacting with the audience; both through advanced humanoid gesticulations and verbalized feedback related to the text. The ability of PiNAOqio to turn the pages of a physical book retains the feel of "real" books that is lost when they are digitized, and also allows users to load their own pre-existing volumes.

PiNAOqio will be versatile enough to be deployed in a range of situations and environments. In particular children that have to be hospitalized for long periods of time experience stress and anxiety [3] and would benefit, having their day-to-day routines greatly enhanced by hearing some of their favorite stories. In no way does PiNAOqio intend to supplant genuine human interaction; it will however serve to make story reading a more fun and social experience for young listeners.

II. HYPOTHESIS

The aim of this project is to investigate the effectiveness of a "storyteller robot" as a source of engagement and educational entertainment for kids. The parameters measured will include user experience and the ability to retain information, upon which the following hypotheses are assessed. These figures have been arbitrarily selected to indicate satisfactory performance:

a) Story-telling by PiNAOqio will improve information retention and user experience by 15% and 30% respectively in comparison with story-telling by an audiobook.

b) Allowing the user to interrupt and interact with the robot, as well as make gestures to emulate a human story-teller, will improve the retention of information and the user experience by 10% and 20% respectively.

III. RELATED WORK

The challenge of designing an interactive robot that reads stories whilst engaging with its users has been tackled in the past. There are several existing robots that focus on either the user experience or the learning aspect of the interaction. For instance, the *Storytelling Companion* is designed by the Personal Robots Group by MIT [4] as a social character, relating to children as a peer, not as a teacher. Through a storytelling game that the robot and child would play together on a tablet, the aim was to improve language and vocabulary skills early on to help them succeed in the future. Another example is *TROBO – the Storytelling Robot* [5], which is a plushy robot that connects with an iPad or iPhone to read a digital storybook. Trobo has been designed to improve the user experience by forming an emotional connection with the robot and the story.

Both of the above examples are dependent on tablets to function – devices that children may also associate with other activities such as playing games or watching videos. Usage of tablets for a storytelling robot do not contribute to solving the problem of long-term overuse. Furthermore, none of the mentioned robots use physical pages – as a result, there is no encouragement for the children to read more physical books.

IV. HUMAN-ROBOT INTERACTION (HRI)

The interaction between the robot and the user has a key role in realizing the hypotheses set above. Given the target audience of young children, the PiNAOqio is designed to be simple yet highly interactive. The operation of the robot is illustrated in the flowchart shown in figure 1.

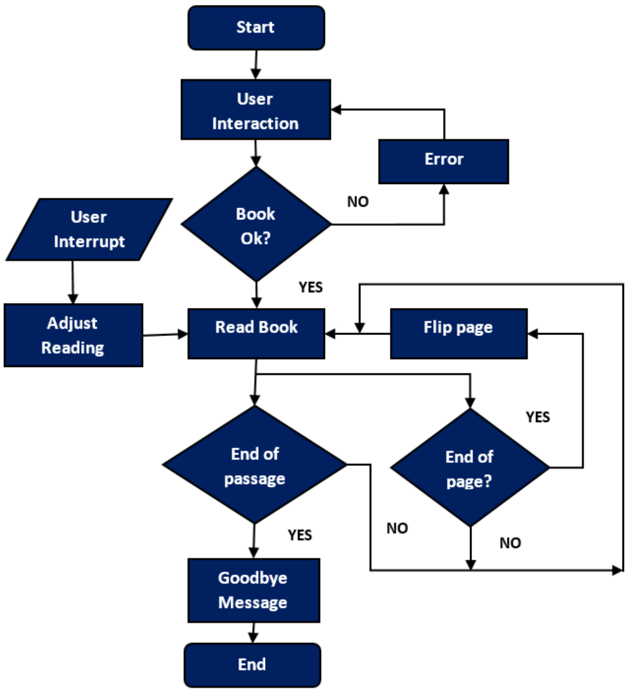

Figure 1: HRI operation flowchart

Initially, there is a short conversation between the robot and the user which acts as an introduction. The aim is to familiarize the user with the robot and its functionalities and perform the loading of the chosen book. If the book has not been appropriately positioned, the robot asks the user to refit it. Once the book is correctly placed, the story begins. In tests where interaction is allowed, the user can interrupt the robot at any time using an external microphone. The user can then instruct the robot to adjust the reading profile to match his preferences by changing the reading speed and tone of voice.

Whilst reading, the robot checks to detect when the end of a page has been reached in order to activate the page flipping mechanism. The page flipping occurs simultaneously as the previous page is being read in order to reduce dead time. Once the entire allocated passage has been completed, the robot will again enter into conversation with the user. The aim is to thank the user for participating in the experiment and give them information about the questionnaires that will be used to assess the retention of information and user experience.

V. HARDWARE

The component to automatically flip the pages of a physical book will necessarily be mechanical and could be approached in a number of ways. Some existing systems accomplish this function using a weak vacuum with enough suction force to lift a page[6]; this can work well and poses little risk of damaging the book, however it is dependent on the suction making good contact with the page. Another approach involves raising a page slightly and using a burst of pressurized air to blow it over[7]; although this method can act the fastest, it is best suited to larger books with heavier pages. Finally, there are many varied systems that use a combination of servos and stepper motors to attain the same end[8][9]. We decided to opt with this latter option as we believed it to be the simplest to construct and also boasts the most reliable operation.

For this implementation to work there are several difficulties that need to be overcome. Firstly, unlike larger textbook types, with small fiction books the pages do not naturally fall to either side. As a result, the mechanism must be able to hold pages in place whilst also allowing for turning. This will require a "holding" servo on either side, where the servo arm can hold down the pages and also move up to catch a new page. Secondly, the curvature of the open book reduces performance of the OCR algorithm. To overcome this a small pressure will need to be applied to hold the book flat. The holding servos should be capable of multifunctioning to achieve this purpose. Finally, throughout the progression of the book the height of front and rear covers will vary. This requires each cover to rest on an independent platform. To deal with this either a stepper motor can control the height of each platform, or each platform can rest on springs where the weight of each side respectively determines height. In figures 2a and 2b below, we can see the page turning mechanism and the sequence of operation respectively.

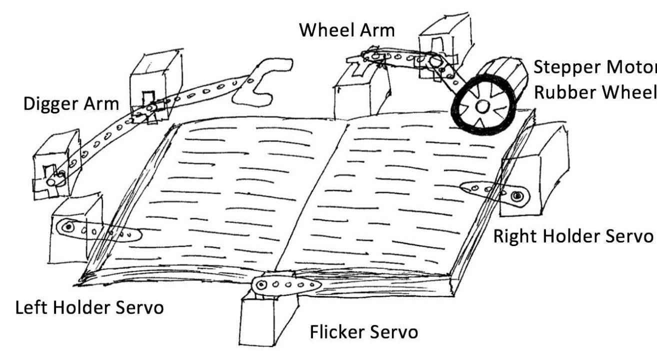

Figure 2a: Page turning mechanism

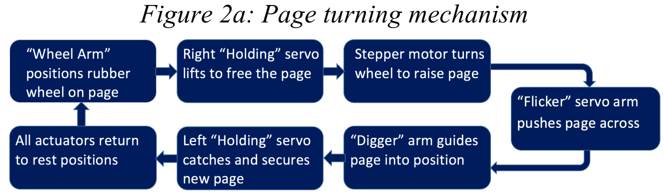

Figure 2b: Sequence of operation

VI. SOFTWARE

We consider the following main software subsystems in our project: the page flipper mechanism, the camera, the gesture movement and the speech. The page turner and camera will be controlled by a Raspberry Pi 3; where as the movement and speech will be controlled by the robot (NAO) hardware. Nao will be using the NAOqi library [10] to communicate with the Raspberry Pi over a wired network connection, which is available on the robot. A server-client model will be employed to facilitate the passage of information between the two subsystems. The Raspberry Pi will behave as a server, waiting for polls from the client (NAO). One advantage of the client-server model in this case is that it gives the client freedom to add extraneous HRI without knowledge of the lower level of the server. The system level of the software operations can be seen in figure 3.

A) Page Turner

A minimum set of high level operations, accessible through NAOqi by a custom API will be implemented in the prototype. Further functionalities may later be added. We assume for the time being that turning a page backwards is redundant for the purpose of simplicity in hardware design. The Raspberry Pi will be responsible for the low-level implementation of each of these functions: Start/Stop/Pause reading and turn page to next.

B) Camera

To 'read' a page, a camera will be used on the Pi to take a picture of the book. Image enhancements will be applied using OpenCV [11] in order to enable robust optical character recognition (OCR) using Tesseract-OCR [12]. Experiments will be performed to determine optimal camera placement and exposure using an artificial source of light which will be present on the final system as well. When a page needs to be flipped, a state machine will sequence the servos as per Figure 2b.

C) Gesture Movement

To enhance the interaction between robot and human, PiNAOqio's motion capabilities are being exploited to provide a natural movement that imitates a human storyteller. NAO's electric motors and actuators are designed and built to allow for 25 degrees of freedom (DOF), which is enough to emulate human gestures. Utilising existing APIs can significantly reduce design and testing time while also improving the user experience. An example includes the *ALAutonomousMoves* API which enables expressive listening through which the NAO makes a slight movement to show that it is listening. A further enhancement will be adding gestures that trigger on specific keywords.

D) Speech

Along with movement gestures, speech processing is an important tool for human-robot interaction. Apart from reading out the text from the chosen passage, speech recognition will be crucial for the initial interaction with the user. These functions will be split between the two following NAOqi APIs: the *ALSpeechRecognition* API, which translates spoken words to text, and *ALTextToSpeech* API, which translates words to speech. The speech recognition will run directly from the NAO as its internal computing resources are sufficient for our purposes. Due to the general complexity of our other hardware and software mechanisms no external speech recognition mechanisms will be used, such as PocketSphinx [13] or Apple's Siri.

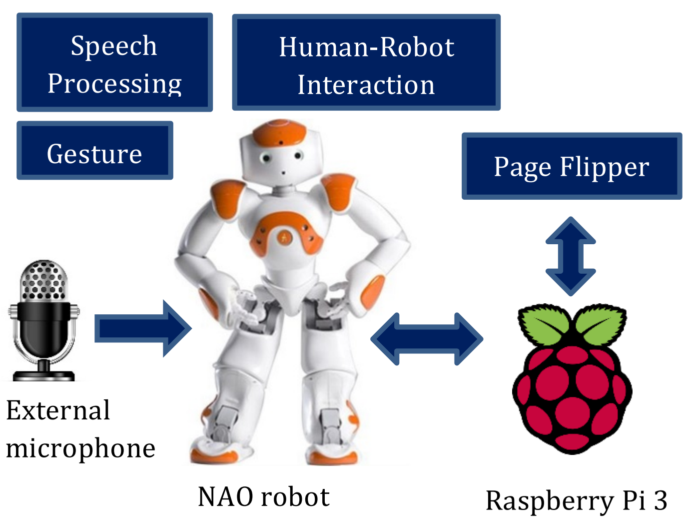

Figure 3: System level view of PiNAOqio

VII. EXPERIMENT

In order to evaluate our hypothesis several tests will be executed. Our experiments will have the objective of evaluating the level of text comprehension of the listeners as well as user experience for different tests: an audiobook, an HRI-deficient robot, and finally PiNAOqio with both gestures and full human-robot interaction functionalities, such as asking for user feedback in real-time or determining passage context.

To facilitate the human-robot interaction through our speech recognition feature, all of our tests will be conducted indoors, in one of the Imperial EEE building rooms. This will avoid a large amount of external noise being added to NAOqi's speech processing tool. Although our objective is to test and use PiNAOqio for children, our initial tests will be done with young adult undergraduate students of all genders from Imperial College London. This is due to our restricted access to children test subjects and ease of availability of Imperial students.

Our examinees will listen to three different passages from the same book, each heard from one of the three different storytellers mentioned above. For each test, we will be providing two different questionnaires: the first analyzing the level of engagement, and the second used to obtain feedback on the user experience. It is important for the responses to be obtained in a quantitative manner in order to build a good comparison between each version being tested. Moreover, we will evaluate the test listener’s engagement level by analyzing their gaze and pupil dilation, through the use of eye recognition technology [14].

VIII. EVALUATION

The performance of PiNAOqio will be evaluated by comparing the user interaction and satisfaction as well as their text comprehension. We will test a fully functioning robot against a reduced interaction system as well as a zero HRI implementation i.e. an audiobook. Comparisons from a text comprehension questionnaire will allow us to understand whether gesticulations and active audio feedback increase engagement and understanding of the text. Using eye recognition technology and cameras that track movement may further enhance the engagement evaluation.

The feedback on the user experience questionnaire will allow us to analyze how the users felt when being read by the two aforementioned instances of our robot with different levels of interaction and the audiobook. Each version of PiNAOqio shall be tested with the same group, in succession, using different passages that convey comparable levels of information.

Multivariate analysis of covariance (MANCOVA) [15] will be used as a statistical tool to determine whether levels of engagement and information retention increase, decrease, or remain unchanged, for a given population over our three storyteller systems. Familiarity with the robot could add a level of dependence across our grouping variable – the different instances – and as such we should threshold our margins for conclusion higher to account for it.

IX. CONCLUSION

PiNAOqio is designed to provide educational entertainment for young people that would otherwise be at risk from the adverse effects of long-term use of electronic devices. It is expected to improve the user experience and text comprehension through human-robot interaction. The novelty in the design arises from the ability to read hardcopy books. Our system will be reading books using external page-flipping hardware and different levels of interaction will be evaluated.

X. REFERENCES

1) Dredge, Stuart. "Children's Reading Shrinking Due To Apps, Games And Youtube". *The Guardian*, 2013. Web. 3 Nov. 2016.

2) Rowan, Cris. "The Impact Of Technology On The Developing Child". The Huffington Post, 2016. Web. 29 Oct. 2016.

3) Hospitalised Childer Stress: W. K. Lau, Bernard and Wilson W. C. Tse. "Psychological Effect Of Physical Illness And Hospitalization On The Child And The Family.", 1993. Web. 1 Nov. 2016.

4) Kory, Jaqueline. "Storytelling Companion - Personal Robots Group". Personal Robots Group, 2016. Web. 28 Oct. 2016.

5) "TROBO The Storytelling Robot: What Are Trobos?". TROBO. N.p., 2016. Web. 30 Oct. 2016.

6) Qumsiyeh, Dany. "Linear Book Scanner - The Open-Source Automatic Book Scanner". Linearbookscanner, 2016. Web. 26 Oct. 2016.

7) "Vision Architecture: High Speed Image Processing". K2.t.u-tokyo.ac.jp, 2016. Web. 27 Oct. 2016.

8) "Page Flipper". Page Flipper, 2016. Web. 27. Oct. 2016.

9) Charlton, John. "Mechanical System For Lifting And Turning Bound Pages". (2015): Web. 28 Oct. 2016.

10) "Naoqi Apis — Aldebaran 2.1.4.13 Documentation". Doc.aldebaran.com, 2016. Web. 28 Oct. 2016.

11) "Opencv | Opencv". Opencv.org, 2016. Web. 29 Oct. 2016.

12) "Tesseract-Ocr". GitHub, 2016. Web. 31 Oct. 2016.

13) "Basic Concepts Of Speech [Cmusphinx Wiki]". Cmusphinx.sourceforge.net, 2016. Web. 28 Oct. 2016.

14) Wang, Joseph. "Pupil Dilatation And Eye Tracking", 2016. Web. 29 Oct. 2016.

15) "Multivariate Analysis Of Covariance (MANCOVA) - Statistics Solutions". Statistics Solutions, 2016. Web. 1 Nov. 2016.