Project Proposal: An interactive picture book reader

COMP3003

Yantao Tang (16522122)

psyyt8@nottingham.ac.uk

October 11, 2019

1 Background and motivation:

Albert Einstein once said: "If you want your children to be intelligent, read them fairy tales. If you want them to be more intelligent, read them more fairy tales". Books not only can help children develop basic language skills [1] but also can develop children's imaginations and expand their world [2]. Early reading lays an important foundation for laying the foundational skills for children, also lays the foundation for their success in school and life. More importantly, parents read to kids is a key component of reading skills development [3].

However, During the average workweek, parents only read 32 minutes with their children [4]. Therefore, a picture book reader that can accompany your child to read is very useful. For example, Luka is an AI picture book reader that reads books through image processing and text recognition [5]. But this product currently only has the ability to read picture books, and cannot interact with children, making this time a little boring. So, it would be good to improve with an interactive feature so that children can ask the AI reading companion for questions about the book without parents. And also, the interactive feature including Automatic Speech Recognition (ASR), Natural Language Processing (NLP) and Text to Speech (TTS) can help kids find their passion for reading.

Andrew Ng once predicted that if Automatic Speech Recognition increased from 95% accurate to 99%, it will become the main way that we interact with computers. Most recently, ASR has benefited from advances in deep learning and big data, and it finally cresting that peak [6]. ASR has invaded our life. Just by speaking, mobiles, games, smart watches and even house can be controlled by voice. Therefore, it is highly feasible to provide an ASR for the picture book reader to get the plain text from the sound wave.

Natural Language Processing is a sub-area of linguistics and artificial intelligence related to the interaction between human language and computer [7], especially indicates the method to program computers to process, analyse and understand the specific content of human language such as questions then provide a highest-ranking candidate answer [8]. With this, the reader is able to reply to the plain text output from ASR in text.

Text to Speech means the computer has the ability to read text aloud. For now, people have solved lots of challenges like using various heuristic techniques to disambiguate homographs or make the speech more prosodics and emotional [9]. Accurate pronunciation and emotional answers will make this reader more vivid, to some extent eliminate the child's strangeness to the reader and make them more interested in it.

Optical Character Recognition (OCR) is able to recognize text from the electronic or mechanical conversion of images including pre-processing to improve the chances of successful recognition, character recognition and post-processing to increase the accuracy [10]. It is widely used in the extracted text-to-speech field. With TTS technique, the reader is allowed to scan the image and read the text aloud.
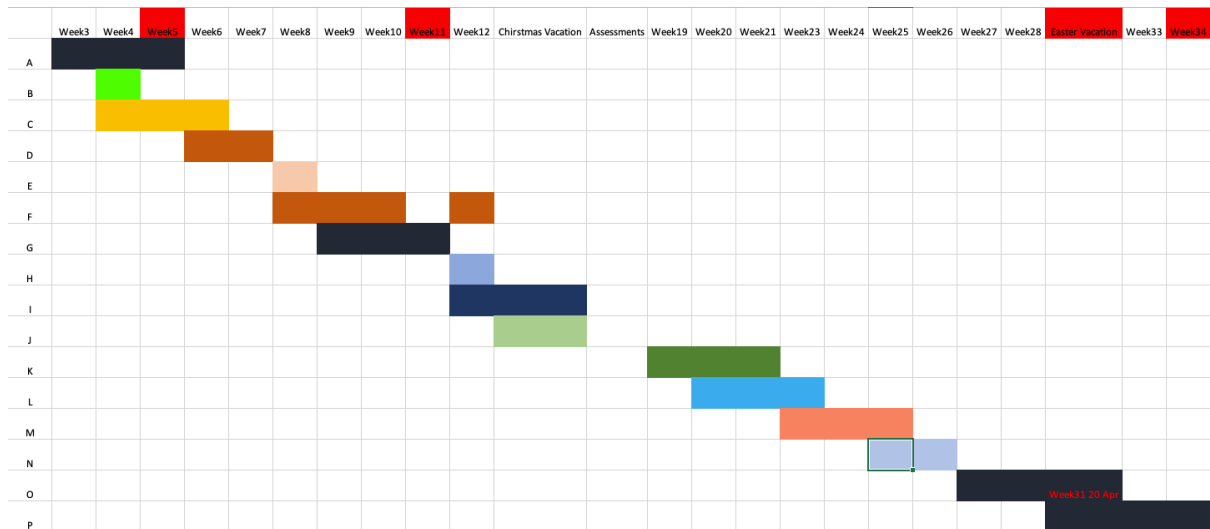
2 Aim(s) and objectives:

The aim in this project is to develop a picture book reader that included Optical Character recognition (OCR), Automatic Speech Recognition (ASR), Natural Language Processing (NLP) and Text to Speech (TTS) so that children could have a basic Q&A with the picture book reader to solve the problem that children need to ask questions about the contents of the book when the parents are unable to accompany them. Also, the AI companion could make reading more fun and is more likely to provoke children's interest in reading.

The key objectives of the project are:

1. Build the ASR feature to transfer the sound wave to the text. It's including extract speech to features, design and train an acoustic model with training labels (characters and phonemes), predict characters with phonemes, dictionary and model and improve by comparing the word error rate.
2. Build the NLP feature to provide a highest-ranking candidate answer to the text transferred. It's including apply semantic role labelling to questions, extract phrase from questions and ranking the candidate answer.
3. Build the TTS feature to speak the content. It's including determine how to disambiguate homographs and decide the voice to ensure it's more vivid.
4. Build the OCR feature to extract text from images. It's including pre-processing such as binarization and segmentation, characters recognition and post-processing like Near-neighbour analysis.
5. Test, evaluate and improve after integrating all four features.
6. Develop a visualisation tool which will allow users to speak and show the results of each three feature directly.

3 Work plan:

A. Complete project proposal and research ethics checklist.
B. Create a GitHub project and create a board on Trello.
C. Research about the decoder of ASR, determine a candidate Neural Networks.
D. Extract speech to features.
E. Find training labels and language model.
F. Build and train the decoder, finish the ASR.
G. Write the interim report.
H. Design the algorithm and choose a framework for NLP.
I. Build and train the NLP feature.
J. Research and design the algorithm for TTS.
K. Build and train the TTS feature.
L. Build the OCR feature.
M. Integrate all four features then test, debug and evaluate.
N. Develop a visualisation tool.
O. Write the final dissertation.
P. Prepare for the final presentation/demonstration.

	Week3	Week4	Week5	Week6	Week7	Week8	Week9	Week10	Week11	Week12	Chirstmas Vacation	Assessments	Week19	Week20	Week21	Week23	Week24	Week25	Week26	Week27	Week28	Easter Vacation	Week33	Week34
A																								
B																								
C																								
D																								
E																								
F																								
G																								
H																								
I																								
J																								
K																								
L																								
M																								
N																								
O																						Week31 20 Apr		
P																								

Reference list:

[1] Kalb, G. R.J., and J. C., "Reading to Young Children: A Head-Start in Life?" SSRN, 20-May-2013. [Online]. Available: https://poseidon01.ssrn.com/delivery.php?ID=78600807300611212400411511106808202600702007 005102903412312302610808809112406412007400509603905906000802707609408208409109611 811701706906902810408211702203108800607005501802307012207911109408106812512007700 602102311609209809608911000100109711608407409&EXT=pdf. [Accessed: 09-Oct-2019].

[2] "Why Do Kids Need Books?" The National Children's Book and Literacy Alliance. [Online]. Available: https://thencbla.org/advocacy/why-do-kids-need-books/. [Accessed: 08-Oct-2019].

[3] D. P. Sukhram and A. Hsu, "Developing Reading Partnerships Between Parents and Children: A Reflection on the Reading Together Program," SpringerLink, 08-Jan-2012. [Online]. Available: https://doi.org/10.1007/s10643-011-0500-y. [Accessed: 10-Oct-2019].

[4] E. Elsworthy, "Half of parents spend less than an hour a day of quality time with their kids," mirror, 29-Jan-2018. [Online]. Available: https://www.mirror.co.uk/news/uk-news/half-parents-spend-less-hour-11933959. [Accessed: 10-Oct-2019].

[5] "Luka AI reading companion," Luka AI reading companion. [Online]. Available: https://luka.ling.ai/. [Accessed: 09-Oct-2019].

[6] "Speech recognition," Wikipedia, 25-Sep-2019. [Online]. Available: https://en.wikipedia.org/wiki/Speech_recognition#End-to-end_automatic_speech_recognition. [Accessed: 10-Oct-2019].

[7] "Natural language processing," Wikipedia, 05-Oct-2019. [Online]. Available: https://en.wikipedia.org/wiki/Natural_language_processing. [Accessed: 11-Oct-2019].

[8] "US8832064B2 - Answer determination for natural language questioning," Google Patents. [Online]. Available: https://patents.google.com/patent/US8832064B2/en. [Accessed: 11-Oct-2019].

[9] "Speech synthesis," Wikipedia, 07-Oct-2019. [Online]. Available: https://en.wikipedia.org/wiki/Speech_synthesis#cite_note-48. [Accessed: 11-Oct-2019].

[10] "Optical character recognition," Wikipedia, 16-Oct-2019. [Online]. Available: https://en.wikipedia.org/wiki/Optical_character_recognition. [Accessed: 16-Oct-2019].