Chen, W. et al. (Eds.) (2016). Proceedings of the 24th International Conference on Computers in Education. India: Asia-Pacific Society for Computers in Education

Automatic Assessment of Reading with Speech Recognition Technology

Preeti RAO*, Prakhar SWARUP, Ankita PASAD, Hitesh TULSIANI and Gargi Ghosh DAS

Department of Electrical Engg., I.I.T. Bombay, India
*prao@ee.iitb.ac.in

Abstract: In this paper, we describe ongoing research towards building an automatic reading assessment system that emulates a human expert in a spoken language learning scenario. Audio recordings of read aloud English stories by children of grades 6-8 are acquired on an available tablet application that facilitates guided oral reading and recording. The created recordings, uploaded to a web-based ratings panel, are currently evaluated by human experts on four relevant dimensions. Observations of typical learner progress patterns will form the bases of a system that applies Automatic Speech Recognition (ASR) techniques to obtain robust automatic predictions of reading fluency and word decoding accuracy.

Keywords: oral reading assessment, tablet application, automatic speech recognition

1. Introduction

It is well known that in India's large rural population, millions of children complete primary school every year without achieving even basic reading standards (ASER, 2012). Since reading competence enhances overall learning by enabling the child to self-learn various subject material from the vast available text resources, the importance of imparting reading skills in early school cannot be over stated. Technology holds the promise of scalable solutions to alleviate the literacy problem. At least one recent effort that has gained visibility is the introduction of a feature known as same-language subtitling (SLS) in which synchronized text subtitles are incorporated in popular TV programs, including songs, so that viewers benefit from exposure to the script while simultaneously listening to the audio (Kothari et al. 2002). A further step in literacy training at the school level would be to facilitate oral reading. Reading aloud has traditionally been an important instructional component in the school system in many countries (Dowhower, 1994). Reading research articles over the decades have presented empirical evidence that assisted oral reading, while simultaneously listening to a fluent reader, is very effective in improving a student's reading skills. Repeated readings of a passage, whether assisted or unassisted, have been shown to lead to improvements both in word decoding and in comprehension. Further, these benefits carry over to new unpracticed texts (Dowhower, 1994; Rasinski, 2003). Given the importance of inculcating oral reading skills, specific scoring rubrics have been developed by educators to evaluate reading fluency in terms of accuracy, rate and expressiveness.

It is the goal of the present work to consider scalable technology solutions that facilitate oral reading practice and assessment in situations where access to language teachers is limited. We choose the specific context of L2 English which is a curriculum subject across schools in rural India where the medium of instruction is primarily the regional language. Qualified English teachers are scarce and even if the books are available, the students' exposure to communicating via speaking is severely limited. We describe a tablet based app that facilitates oral reading practice and present research that targets the automatic assessment of recorded speech according to accepted norms of reading proficiency. The goal is to achieve reliable means of objective feedback to (i) the student as a motivational component, and (ii) higher authorities in the school system as a meaningful student progress tracking. The proposed solution, developed for English, can be extended to any language.

Automatic speech recognition (ASR) is a technology that converts an acoustic speech signal to text using language and other constraints. It can therefore be applied, in principle, to evaluate the accuracy

of the read speech of a student with reference to the known text. Further speech analysis techniques are available to derive other aspects of speech delivery such as expression. However, ASR is most successful in applications where the variability in the acoustics due to speaker and environment conditions are controlled. In the school reading scenario, we expect significant diversity in speaking accents and pronunciation, and the ubiquitous presence of background noise. To obtain robust ASR in this case, we need a system that is built specifically for the task and anticipated use case. A thorough understanding of the feedback and evaluation requirements and of the type of variations that occur in practice will be useful in addressing the mentioned challenges.

In the next section, we describe the tablet application we employ that serves to facilitate oral reading and recording. This is followed by a discussion on known attributes of oral reading proficiency which serve to define the scope of the automatic assessment. Methodologies for collecting and labeling audio data to train the machine learning algorithms of the ASR system are discussed.

2. The tablet reading application

Mobile tablets provide for a low cost, portable devices that can be easily handled by children. The screen space of a 7 inch tablet is sufficient for the convenient display of text and pictures in story reading. Interactivity can be easily incorporated via touch. Additionally, a microphone and camera are always available on device. Sufficient memory and internet connectivity ensure that the tablet can be embedded in a larger connected system where content delivery and transfer of data are easily achieved. There are a number of Android OS based tablets in the market that satisfy these basic requirements. There are a few apps available for such tablets that facilitate guided reading accompanied by a narrator voice and the display of text with word-by-word highlighting at a suitable reading pace. For the proposed work, we adopt the SensiBol Reading Tutor app (2016) for Android tablets due to the availability of customization for classroom use with multiple separate child accounts. The app allows a child to listen to a narrator reading out the story in a listening mode. The child can use the record mode while reading aloud herself. Both unassisted and assisted (i.e. shadowing the narrator) modes are available in the read aloud mode. The stored child recording synchronized with the video is available on the tablet for listening which feature encourages self-assessment and more practice. In order to make it even more interesting, the child's recording can be enhanced by adding audio effects and by mixing in any background track available in the original story video. Further, the child recording is also available to the teacher for review at any time. All recordings are made with a headset mic to minimize background noise which can be very deleterious for ASR. The content is a selection of stories from BookBox (2016), a readily available rich resource of illustrated text designed for child readers.

The SensiBol RT app also provides backend support where every registered child's audio recordings, together with metadata information such as child name, story name, date and time can be archived. A web-based ratings panel displays the audio at the sentence level together with the expected story text. This is obtained by segmenting the full story recording based on a combination of information from the video timings combined with the detection of long silences in the audio. The sentence-level audio can then be rated by a human expert on various dimensions in comparison with the corresponding narrator audio. This facility is vital for the process of creation of the labeled data resources required for ASR based system development. The labeling exercise is underway based on the field testing deployment of the reading app in a tribal school near Mumbai involving children in grades 6-8 where tablet based story reading is a scheduled and supervised activity conducted in the school hours (LETS, 2016).

3. Automatic prediction of reading skill with ASR

Proficient readers organize the text into meaningful phrases and read with appropriate prosody and pace apart from the correct pronunciation of words. Comprehension has been shown to be predictable by the prosody of the student's reading, i.e. its pauses and intonation (Miller, 2008). Thus prosodic oral reading can signal that children have achieved fluency and are more capable of understanding what they read. This suggests that performance in different dimensions, word decoding and prosody related, must be considered by an automatic assessment scheme. Our observations on field

recordings of repeated readings (interspersed with listening to the narrator) of a single story collected from a group of 5 children (in grades 6-8) over a span of several days indicate the following:

1. The child typically starts out with reading out word by word in list style including disfluencies comprising of long pauses, hesitations and incomplete words.
2. Improvements are always noted in one or more dimensions from reading to reading (where there is a separation of at least one day). The earliest to improve is phrasing (pauses reduce, words become more smoothly connected). The reading pace gradually increases.
3. After 2 or 3 readings, the volume and intonation start varying (this also conveys improved confidence). Eventually, the correct intonation, matching that of the narrator, is achieved.
4. Some word pronunciations get gradually clearer. But certain wrong pronunciations, especially those linked to grapheme-to-phoneme confusions and L1 influence, persist.

We can conclude from the above that evaluation should target phrasing and reading pace in the first instance, followed by prosody. Finally, fine feedback on word pronunciations can be provided. Thus the level of feedback is adapted to the child's proficiency level. In order to build training datasets for the different stages of machine based evaluation, we are in the process of obtaining expert ratings of word accuracy (word substitution, insertion and deletion) and three prosodic attributes, viz. phrasing, volume and intonation and reading pace, on a 4-level scale at the sentence level. These ratings are obtained for each of several repeated readings of 18 distinct stories by 100 children from the selected school and grades. The word level transcriptions will be used to build acoustic models and language models for the ASR system that incorporate typical pronunciation variations and observed disfluencies. The prosody ratings will be used to train classifiers on selected acoustic features to detect proper phrasing and sentence intonation. All this functionality further needs to be robust to at least the low levels of background noise and interference expected in the audio.

While ASR has been used previously in objective assessment of language skills of children, the present work is targeted towards the more challenging scenario of continuous speech (rather than isolated words as in the extensive work by Alwan et al. (2007)). Achieving robust automatic evaluation will also help us to introduce more sophisticated testing for comprehension such as story retelling exercises.

Acknowledgements

We would like to thank Nagesh Nayak and Sujeet Kini of SensiBol Audio Technologies for help with the customization of the Reading Tutor App for this study. We thank Prof. Alka Hingorani and her team at IDC, I.I.T. Bombay for initiating the LETS project and organizing the field study.

References

ASER, The Annual Status of Education Report (Rural) (2012) url: http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER_2012/fullaser2012report.pdf

Kothari, B., Takeda, J., Joshi, A., & Pandey, A. (2002). Same language subtitling: a butterfly for literacy?. *International Journal of Lifelong Education*,*21*(1), 55-66.

Dowhower, S.L. (1994). Repeated reading revisited: Research into practice. *Reading and Writing Quarterly*, 10, 343-358.

Rasinski, T. V., & Hoffman, J. V. (2003). Oral reading in the school literacy curriculum. *Reading Research Quarterly*, *38*(4), 510-522.

SensiBol Reading Tutor App, SensiBol Audio Technologies Pvt. Ltd. (2016), url: http://sensibol.com/readingtutor.html

BookBox, A book for every child in her language. (2016) url: www.bookbox.com

LETS (Learn English Through Stories), Tata Centre for Technology and Design at I.I.T. Bombay (2016). url: http://www.tatacentre.iitb.ac.in/15mobitech.php

Miller, J., & Schwanenflugel, P. J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly*, *43*(4).

Alwan, A., Bai, Y., Black, M., Casey, L., Gerosa, M., Heritage, M., & Narayanan, S. (2007). A system for technology based assessment of language and literacy in young children: the role of multiple information sources. In *Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop*.