Automatic assessment of children's oral reading for prosodic fluency

Kamini Sabu

Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India

kaminisabu@ee.iitb.ac.in

Abstract

Reading skills are essential for better understanding and effective communication. Low literacy skills all over the world can be mainly attributed to unavailability of good teachers. In this scenario, we aim to build an automatic reading evaluation system to assess children's oral reading in terms of reading miscues, speech rate and fluency as per the NAEP reading scale. We aim at building a text dependent reading assessment system using automatic speech recognition and prosody modeling. The performance is evaluated on English story recordings of oral reading by native Marathi students of age group 10-14 years in India. Students are divided into two groups with good and poor word decoding ability. Phrasal breaks and prominent words are detected in the recordings. We propose to use these along with other prosodic features for high-level ratings corresponding to confidence, expressiveness and comprehension.

Index Terms: oral reading assessment, prosody, expressiveness, reading skill evaluation

1. Introduction

Reading skills are essential for better understanding and comprehension, effective communication and building confidence for speaking. Reading difficulty, on the other hand, contributes to school failure increasing the risk of absenteeism further raising percentage of illiteracy. As per the Annual Survey Education Reports [1] by Pratham over last 4-5 years, basic literacy skills of children in rural India are below par. Almost 50% of the students can't read correctly and fluently, English as second language being even more susceptible. Even the 'education for all' movement by government does not seem to improve the reading skills of children. Important causes behind this fact include shortage of qualified teachers and dismal student-teacher ratio.

In this scenario, we aim to build an automatic reading evaluation system to assess children's oral reading in terms of reading miscues, speech rate and fluency as suggested by the NAEP reading scale [2]. We further aim to assess overall reading ability specifically inclined towards estimating comprehension/understanding level. Literature claims that reading prosody is a clear indicator of comprehension and teachers are advised to use realization of phrase boundaries and expression as important criteria towards the assessment of oral reading skills [3, 4]. With this, we propose to use student's reading prosody for assessing the expressiveness and comprehension.

1.1. Related Work

Most of the previous work on automatic reading tutors focuses on pronunciation accuracy and lexical miscues only. There is limited previous research on exploiting prosodic cues to evaluate reading skills. Among the works using prosody for reading evaluation, some works refer the use of gold standard from good readers students, adults or professionals [5, 6]. These standard recordings are used as reference and compared with the test readings to give the assessment scores. However, these systems face issues while adding new reading material for which standard voices are not available. Besides, students are known to make many mistakes while reading which include omissions, insertions and substitutions along with disfluencies, hesitations, etc. These are not included in the standard recordings, and hence the direct comparison between standard voice and test reading may not be possible.

1.2. Thesis Proposal

We aim at building a text dependent assessment system similar to [7], which uses different lexical and prosodic features in an SVM classifier. This system, however, estimates a single score corresponding to the overall literacy skills of students, while we propose to build an assessment system to rate the oral reading skills at granular level. This may help give proper feedback for further improvement. Students gradually progress through different stages towards attaining good reading skills. These stages can be sequentially listed as good word decoding ability, increased reading speed, bigger and meaningful chunk formation, and emphasis on new/important words [8, 9]. Accordingly, we propose to assess reading in terms of four attributes, viz. word decoding ability, speech rate, phrasing and prominence, similar to MDF scale proposed by [10].

We use Automatic Speech Recognition (ASR) module to estimate the word decoding ability and speech rate. It has been observed that beginning readers read in monotonous voice, with wrong intonations and introduce pauses at wrong positions compared to practiced readers. This is clearly reflected through pitch contour trends, intensity variations and pause structure (duration and position). These features can also be used to detect lexical disfluency [11] and hence to separate children with poor word decoding ability from those with good word decoding ability [12, 13]. Further evaluation can then be performed for the later group only. Since the story context is known, a good comprehending reader is supposed to be able to determine intonation as per the syntax and important words as per the information structure. Phrasal break positions and prominent word positions can be detected using prosodic features like duration, pitch variation and energy [14, 15, 16]. We then propose to use these detected prosodic event positions for scoring the comprehension ability. Further, fixed rhythm or sing-song style indicates comparatively reduced understanding than when certain intended words are stressed. Similarly, the expressiveness and confidence level can also be assessed using prosodic features, e.g. pitch and intensity variation is indicative of expressiveness [17] while pitch and vocal quality contribute to confidence [18].

2. Dataset and Annotation

Dataset for this work is collected from native Marathi students of age group 10-14 years in urban schools in India. Short En-

glish stories are presented to them on printed sheet of paper and an audio recorder app on a tablet is used for recording using a headset microphone. Data is specifically collected from children studying in grades 5-7 in an urban school in perceptually clean surroundings. Students are selected such that they have good word decoding ability, but one can observe wide variability in their prosodic skills. The data is transcribed and every word labeled with lexical miscue tags. Commonly observed disfluency types like hesitation, sound outs and mumbling are also labeled separately. Prosodic events labeling is done by naive listeners who label every word as 'prominent' or 'non-prominent' and every word boundary as 'phrasal break', 'sentence break' or 'no break'. Inter-rater agreement is measured in terms of Fleiss' kappa which is 0.4 for prominence and 0.7 for phrasing for this data. The ground truth for the higher level attributes corresponding to comprehension, expressiveness and confidence are obtained from language experts and/or teachers. For this, raters are given part of recordings equivalent to 4-6 sentences in the story. Raters then score the attributes on 3-point scale. The assessment system is supposed to mimic the same.

3. System Description

The overall proposed system is as shown in Figure 1. Test

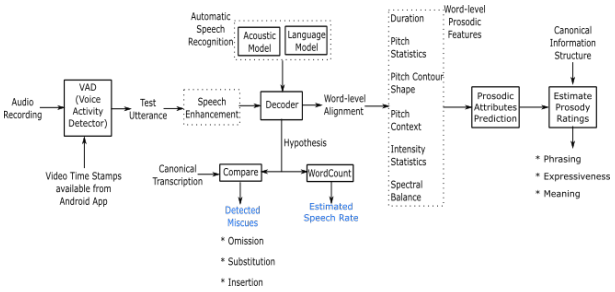

Figure 1: *Overall System Block Diagram*

recording is passed through a voice activity detector (VAD) to remove long noisy/silence regions which may be otherwise responsible for failure of acoustic model in ASR. The test utterance is then enhanced for noise suppression, if needed. ASR decoder yields corresponding text hypothesis and phone- and word-level alignments. The ASR decoder comprises of a state-of-the-art acoustic model coupled with a text-specific language model for miscue detection and alignment estimation. The language model is story specific trigram with a generic garbage model. Garbage model is universal with all possible variations in all the words - valid or invalid - added manually. The text hypothesis is compared with canonical text to estimate the speech rate and yield detected miscues in the form of omission, substitution, etc. Word-level alignments are used to get word-level prosodic features related to duration, F0, energy and spectral shape. The prosodic features are fed to the prosodic attribute prediction classifier to detect prosodic events in the recording. Prosody rating estimates are then obtained in the form of phrasing, expressiveness and meaning.

We calculate prosodic features from pitch and intensity contours and feed these to a random forest classifier to predict prominence. With the previous study [19] on determining features important for implementing prominence, we found that syllable duration, pitch span, likelihood of pitch contour to peaking shape and spectral tilt span are important cues indicating prominence. Accordingly, we have selected a set of 40 features which are used in prominence detection classifier. We also tried word-level features for phrase break detection [20], but it seems to have less precision-recall. Instead, it seems that phrasing can be more accurately predicted with features related to pause structure and pitch reset characteristics [21].

4. Results Discussion and Future Plans

From the results so far, we observed that the developed system predicts prosodic events with almost 75-80% accuracy. The cases where it fails include wrong word-alignments, inaccurate pitch contour, etc. Another important concern is the interdependence of prosody attributes. e.g. syllable lengthening and pitch decline on phrase-final word are indicative of good phrasing. The same (duration elongation and high pitch span) also form cues to prominent words leading to false detection of phrase-final words as prominent.

The target class for the work is children. Children are by default known to be highly inconsistent leading to difficulties in modeling their speech. Regional native language for the children under consideration is Marathi, English being second language. This introduces native language accent on their English reading. Since they are new learners of the language, hesitations and filled-pauses are often observed in recordings. Because of the confusing use of prosodic features by children, inter-rater agreement among experts is also very low.

With this, the future work is expected to comprise the investigation of different techniques for prosody modeling to assess fluency. Different ways of normalization and adding context window can be investigated for better performance of prosodic event classifiers. Prosodic features in the context of regional accent may help further. ASR metrics like word confidence score can be implemented to detect lexical disfluencies along with the number of miscues and speech rate. Good ASR, and prosodic feature extraction techniques can be devised for improved performance. We need to devise a better way of capturing ground truth for prosodic event labeling which will help further estimation of final high-level ratings. This may include discarding the ambiguously rated entities. Next step towards achieving the final goal is to devise classifiers with different prosodic features and detected positions of phrasal breaks and prominent words for estimating the comprehension. Approaches to estimate other high level ratings like confidence and expressiveness are to be investigated. Final goal of the project is to rate students' reading skills which will help students identify areas to improve on and teachers to identify deficiencies of every child and channel their efforts accordingly. Text independent system, based solely on prosody and not using ASR at all, neither for word alignment nor for lexical disfluency, can also be tried. In this case, we may evaluate solely based on some energy and pitch based speech continuity and rhythmic measures.

Extending the same automatic assessment system for reading assessment in different Indian languages can also be considered if time permits. Further, in order to make the system practically usable in general scenario, it should be robust to noise in school surroundings. Good speech enhancement system may be helpful in this regard.

5. Acknowledgements

I would like to thank my supervisor Prof. Preeti Rao for her invaluable guidance and support and my labmates for their help in system implementation. I acknowledge funding support from Visvesvaraya PhD scheme by Ministry of Elect. and IT of India.

6. References

[1] "ASER: The Annual Status of Education Report (Rural India)," http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER%202016/aser_2016.pdf, ASER Centre, 2016.

[2] N. R. Panel, "Teaching children to read:an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction," The Eunice Kennedy Shriver National Institute of Child Health and Human Development, Tech. Rep., 2000.

[3] M. Breen, L. Kaswer, J. V. Dyke, J. Krivokapic, and N. Landi, "Imitated prosodic fluency predicts reading comprehension ability in good and poor high school readers," *Frontiers of Psychology*, vol. 7, pp. 1–17, 2016.

[4] R. Hudson, P. Pullen, H. Lane, and J. Torgesen, "The complex nature of reading fluency: A multidimensional view," *Reading & Writing Quarterly*, vol. 25, no. 1, pp. 4–32, 2008.

[5] P. Black, J. Tepperman, and S. Narayanan, "Automatic prediction of children's reading ability for high-level literacy assessment," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 1015–1028, 2011.

[6] M. Duong, J. Mostow, and S. Sitaram, "Two methods for assessing oral reading prosody," *ACM Transactions on Speech Language Processing*, vol. 7, no. 4, pp. 14.1–14.22, 2011.

[7] D. Bolanos, R. Cole, W. Ward, G. Tindal, P. Schwanenflugel, and M. Kuhn, "Automatic assessment of expressive oral reading," *Speech Communication*, vol. 55, no. 2, pp. 221–236, 2013.

[8] "Why prosody matters: The importance of reading aloud with expression (2018)," http://www.scilearn.com/blog/prosody-matters-reading-aloud-with-expression, Scientific Learning Fast ForWord.

[9] P. Rao, P. Swarup, A. Pasad, H. Tulsiani, and G. Das, "Automatic assessment of reading with speech recognition technology," in *Proceedings of International Conference on Computers in Education*, Mumbai, India, 2016.

[10] J. Zutell and T. Rasinski, "Training teachers to attend to their students' oral reading fluency," *Theory into Practice*, vol. 30, no. 3, pp. 211–217, 1991.

[11] E. Shriberg, R. Bates, and A. Stolcke, "A prosody-only decision-tree model for disfluency detection," in *Proceedings of EUROSPEECH*, Rhodes, Greece, 1997.

[12] J. Liscombe, "Prosody and speaker state: Paralinguistics, pragmatics, and proficiency," Ph.D. dissertation, Columbia University, 2007.

[13] J. Miller and P. Schwanenflugel, "A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children," *Reading Research Quarterly*, vol. 43, no. 4, pp. 336–354, 2008.

[14] S. Ananthakrishnan and S. Narayanan, "Automatic prosodic event detection using acoustic, lexical, and syntactic evidence," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 216–228, 2008.

[15] A. Rosenberg, "Automatic detection and classification of prosodic events," Ph.D. dissertation, Columbia University, 2009.

[16] G. Christodoulides and M. Avanzi, "An evaluation of machine learning methods for prominence detection in french," in *Proceedings of INTERSPEECH*, Singapore, 2014.

[17] D. Bolanos, R. Cole, W. Ward, E. Borts, and E. Svirsky, "Flora: Fluent oral reading assessment of childrens speech," *ACM Transactions on Speech Language Processing*, vol. 7, no. 4, pp. 16:1–16:19, 2011.

[18] X. Jiang and M. Pell, "Predicting confidence and doubt in accented speakers: Human perception and machine learning experiments," in *Procceedings of Speech Prosody*, 2018.

[19] K. Sabu and P. Rao, "Detection of prominent words in oral reading by children," in *Proceedings of Speech Prosody*, Poznan, Poland, 2018.

[20] ——, "Automatic assessment of children's oral reading using speech recognition and prosody modeling," in *CSI Transactions on ICT*, vol. S.I. Visvesvaraya, 2018, pp. 1–5.

[21] K. Sabu, P. Swarup, H. Tulsiani, and P. Rao, "Automatic assessment of children's L2 reading for accuracy and fluency," in *Proceedings of SLaTE*, Stockholm, Sweden, 2017.