Oral Reading Prosody Improvements Using a Reading Tutor for ESL Students

Kalika Bali

kalikab@microsoft.com

Microsoft Research Labs, Bangalore, India

Abstract

Most evaluation of reading assessment of L2 learners concentrates on accurate decoding and ignores objective assessment of prosody. However, fluent reading of a language is often associated with use of expressive and appropriate prosody. This paper looks at data from a 2 ½-month pilot study of Project Listen's PC-based Reading Tutor program at low-income elementary schools in India to investigate the role of prosodic chunking at inter and intra sentential levels in reading fluency. We show that even though there is no clear correlation between the text-based measures and prosodic gains, for the more fluent reader oral reading prosody improvements do occur with the use of the Reading Tutor.

Index Terms: Second language acquisition, Prosody, Reading Fluency, English as a Second Language, Educational Technology.

Introduction

Fluent reading of a language is often associated with fewer pauses, chunking of words into relevant phrases and appropriate intonation [Schwanenflugel, P. J., & Ruston, H. P. (2008)]. While researchers in education psychology and pedagogy might differ in conceptually defining reading fluency, there is an agreement that fluent readers are in general better at comprehension of a text, both in terms of better as well as faster understanding of the meaning of the text [Schwanenflugel & Ruston (2008), Wolf & Katzir-Cohen (2001), Kuhn & Stahl (2004), Kuhn & Schwanenflugel, (2008), Yamashita & Ichikawa 2010]. Kuhn & Stahl (2004) point to three main characteristics of a fluent reader: a) accurate decoding, b) automatic word recognition, and c) use of appropriate prosody. According to them, the comprehension of text at a suitable rate involves skills in all three areas.

Even though "Prosodic Reading" or reading with appropriate prosodic chunking and expression is considered an important stage of fluent reading, it still remains difficult to assess and quantify. Most evaluation of reading assessment concentrate on decoding with measures like word error rates, and tends to use subjective scalar measures for prosody, if at all. There have been significant efforts towards assessing prosody in reading [Schwanenflugel et al (2006), Mostow & Duong (2009), Duong & Mostow (2009)] that clearly indicate that there are measurable positive gains in the expressiveness of a child's reading that can be correlated with quantitative prosodic cues. These metrics maybe manually or automatically determined, and usually compare a child's prosodic contours correlate with an adult's reading of the same text.

However, all the above work concentrates on fluency for L1 learners (mainly English) and there has been little work on assessing prosodic gains for L2 readers. Palumbo & Willcutt, (2006) assert that the link between an expressive reading and comprehension is not straightforward for L2 readers as usually L2 readers start learning to read much before they have the necessary grammar and vocabulary for L2 spoken language. There is a general consensus in L2 reading literature [Yamashita & Ichikawa, (2010), Palumbo & Willcutt, (2006)] that chunking or grouping of

words into meaningful units is fundamental for fluent reading. Although, Yamashita & Ichikawa, (2010) show that the relationship between chunking and comprehension may be more complex, it is indeed a crucial step towards fluent reading and hence, better comprehension.

In this paper, we look at reading data from a 2 ½-month pilot study of Project Listen's PC-based Reading Tutor program [Mostow and Aist, (2001)] for enhancing English education in India. The focus of this pilot was on low-income elementary school students, a population that has little or no exposure to English outside of school but was learning English as Second Language (ESL) at school. The participating students showed a clear improvement in reading fluency based on text-based measures [Weber and Bali, 2010] however, no evaluation was done on their oral reading prosody. This study is an initial investigation into a student data from the pilot with the aim of understanding the effects, if any, of using a reading tutor on oral reading prosody. In the next section we discuss the importance of English education in India and its many manifestations. Section 3 gives a brief overview of the CMU's Project Listen Reading Tutor. The role of prosodic assessment in reading fluency is looked is described in Section 4. The results are presented in Section 5 and Section 6 discusses the implications of these for future work.

1 English in India

English in post-independence India remains an important means for upward mobility. It is recognized in the Indian Constitution as the secondary official language of the country, after the primary official language Hindi. In a country with 22 official languages English remains the language of higher education, business and professional activities.

For most learners of English as a second language, the main driving force to learn English lies in its perception as an enhancer of social status. The mainly instrumental motive [Sheorey (2006), Agnihotri and Khanna (1995)] is also due to the fact that a fluency in English leads to a wider range of employment opportunities, from working in the stores of the upscale shopping malls to serving to differentiate between applicants for basic entry-level employment in business. A good command of English is automatically associated with competence, good education, and intelligence.

This emphasis on English language skills puts the large majority of children going to schools where instruction is in their local language at a distinct disadvantage. Thus many aspirational parents make extensive sacrifices to obtain an English education for their children in private schools. However, a serious shortage of trained teachers makes consistent quality of English teaching in India an uphill task, especially for schools with limited resources which mainly cater to the lower socio-economic classes. Only a handful of schools can provide anything but a minimal of individual attention and practice, a direct consequence of large class sizes and a focus on completing the prescribed syllabus. A strong emphasis on the end of the year examination leads instructors to target written skills at the cost of comprehension and spoken skills [Ramanathan (2008)]. It is precisely these skills, however, especially self-expression in spoken English that are deemed necessary for upward mobility.

Weber and Bali (2010) discuss the potential for introducing technology to improve the reach and effectiveness of English language teaching in the resource-limited Indian context. Their survey of English education programs in and around Bangalore focus on four themes: the demographics served, importance of English literacy in the curriculum, computing and personnel infrastructure, and attitudes towards technological interventions. [Weber and Bali (2010) argue that a computer-based reading tutor can address the issues related to shortage of teachers, and lack of enough time to practice spoken English by enabling an individual to practice spoken skills consistently across all students. It can also assist and test comprehension and spoken fluency. However, they believe that the success of such a technology is dependent on personnel and technological infrastructure, as well as an explicit measure of success in terms of gains in comprehension and spoken skills. "Many of the schools are willing to introduce such software on top of their curriculum if they see it bring value in terms of better performance" [Weber and Bali (2010)].

2 Reading Tutor Pilot

Project Listen's reading tutor [Mostow and Aist, (2001)] provides reading practice targeted at each student's individual reading level through story reading. This PC-based software provides each student with a unique ID, and a choice of stories according to their reading level. On choosing a story, the student is shown the story a sentence at a time and are expected to read it into a headset microphone. A semi-constrained speech recognition engine based on CMU Sphinx is employed to "listen" to the student's speech. The ASR engine keeps track of what the student says and their progress through the story, word by word, sentence by sentence. The program provides feedback whenever a student gets stuck or mispronounces a particular word through a number of strategies: a word-specific prompt that may include similar sounding words that the student might be familiar with and even playback the entire sentence from pre-recordings of the story in a US-accented native voice. Students can also click on difficult or unfamiliar words to get word-specific help.

The students are expected to use the tutor for 30 minutes a day and all interactions between the tutor and each student are recorded in a central database. The tutor and the teachers can monitor the progress of each student individually through the central database. The program uses speaking rate to track students and advance them through increasing levels of complexity. Simple word games are employed at the beginning to determine whether the students have mastered letter-to-sound rules or not. The tutor then proceeds through different levels of new stories from primary level to roughly 8th grade levels of expected fluency. An interesting point to note is that that the reading tutor measures proficiency in reading only through reading rate, no other assessment either through comprehension testing or prosodic evaluation is used to grade the students' progress.

As observed in Weber and Bali (2010) the reading tutor can help address some of the problems faced by an Indian elementary or secondary school English program. It allows an individual student to practice reading in a forgiving environment. Most classroom reading is sporadic, rarely longer than a few minutes at a time and often accompanied by performance anxiety arising from the fact that all your classmates are watching. Also, for many students the reading tutor's use of native English speaking voices for narration, and instruction is their only exposure to fluent, native English speech. The fact that the students can choose the stories they want to read and play games makes the experience less tedious.

The pilot [Weber and Bali (2010)] was run at 3 schools (S1-S3) chosen after a detailed survey to target the underprivileged demographics from low-resource government as well as private schools. S1 had 9 students from Grade 9, S2 22 students from Grade 4, and S3 had 30 students studying in Grade 3. The students at all three sites were divided into two groups, the first group had additional half an hour a day 5 days a week practice with the reading tutor, while the second group had no extra reading other than the regular classroom activity for the entire class. Mid-way through the project the two groups swapped and the second group now worked an extra 30 minutes a day with the reading tutor while the first group was restricted to only classroom reading. All time with the reading tutor was in addition to the time spent in regular classroom instruction. Both the groups were tested (see next section) at the start, middle and end of the pilot to evaluate gains made with and without the use of the tutor. While there was considerable variation in individual and group results across all sites and grades, the pilot was able to show significant positive gains through the use of the reading tutor.

3 Prosodic Assessment of Reading Fluency

As mentioned in Section 1, fluent reading is strongly associated with the use of appropriate prosody. An assessment of reading fluency especially oral reading should incorporate metrics for prosodic gains. However, there is some debate on which prosodic features can jointly represent "appropriate" prosody for fluent reading (Schwanenflugel et al (2006)). Further, some research [Cruttenden (1985), Cutler & Swinney (1987)] argues that it is not clear that young children have sufficiently developed decoding skills to be able to generate prosodic reading of a text. While it may not be clear how

prosody fits in the broad framework of reading skills there is enough evidence that improvement in reading skills is correlated to expressive oral reading.

Schwanenflugel et al (2006) in a fairly comprehensive study that aims to characterize the development of reading prosody as a function of decoding skills compared prosodic features of primary grade children's reading with those of fluent adult readers. They analysed the speech waveforms and f0 contours across five measures, viz., a) intersentential pause length means, b) intersentential pause length variance, d) child-adult f0 profile match, and e) sentence final declination for declarative sentences. Their study showed that primary grade children with fairly developed decoding skills had fewer and briefer intrasentential pauses, and a good match with the f0 profile of adults. These children had clear end of sentence (EOS) markings using f0 declination more than intersentential pauses. In contrast, the children with lower decoding skills read slowly with numerous long pauses and a flat f0 contour. While there results failed to show any clear link between any improvements in comprehension, they do characterize the development of reading prosody as a function of decoding skills.

Mostow & Duong (2009) present a method for automatic assessment of oral reading prosody, and its evaluation on a large corpus of real data recorded by students using Project LISTEN's Reading Tutor at school. They tested their assessment both against human scores, and indirectly by predicting scores and gains on standard tests of oral reading fluency and comprehension. Although their method did not succeed in estimating fluency scores for individual sentences, it did better at estimating scores at the student level, and out-performed other methods in predicting students' fluency posttest scores. It predicted gains in fluency and comprehension much better than standard pretest scores did. In a related study [Duong & Mostow, J. (2009)] they found that the prosodic features related to matching adult f0 contours with the children are less sensitive in predicting the gains in fluency compared to those based on individual student's speech. They also show that statistically significant recency effects are much stronger than learning effects, that is, children show more improvement on rereading text on the same day than several days later. Both their studies are based on several thousand sentences over more than a hundred students.

In Weber and Bali (2010) two test regimes were used to quantify progress of the students using the reading tutor. The first followed Curriculum-Based Measurement (CBM) [Deno (1985)] where each student was made to read out a passage selected from a story at the student's reading level that they had not seen before. Each child read from the passage for a minute and was scored on the numbers of words read correctly, without pronunciation errors or hesitations. The second Test of Written English (TWE) is a commercially developed spelling test [Larsen et al (2009)] that involves presenting a series of spelling words of increasing difficulty to each child; the score is the number of words correct before five consecutive errors. While it was not clear whether TWE was sensitive to any gains made during this relatively short pilot, CBM, though a very coarse grained unsophisticated rubric, was able to capture the gains made by the students to a large extent. Admittedly, none of these scores attempted to capture any prosodic cues for fluent reading.

Sitaram and Mostow (2012) in mine a large database of children's reading data to analyse textual and syntactic features that best predict the reading gains. Their analysis shows that the adult prosodic patterns can differ in various contexts from children's. Also, they show that while prosodic features might be indicative of reading fluency there may be other linguistic factors that need to be taken into account. However, as they themselves point out, the data for their study uses a single sentence read out from the screen by the users and does not account for "expressive fluency" of connected speech.

4 Results

In our assessment of prosodic reading we looked at the recordings of the tests conducted during Weber and Bali (2010). We divided the data into two groups based on their scores in CBM based evaluation. Group A included students from across the three sites with a change in CBM score of greater than 7 and Group B included

Table 1 (a): *The average number of Intrasentential pauses*

students with change in CBM score of less than 7. We chose 30 sentences per student for evaluation. A number of potential Group B students had to be excluded from the evaluation as they had failed to read out even one sentence in a minute for both pre and post testing. A total of 12 students in Group B and 13 in Group A were considered for evaluation. It is important to point out that none of the participating students could be termed a "fluent reader of English", even then, as can be seen in the results presented, they showed significant improvements in their reading skills.

A modified sub-set of measures presented in Schwanenflugel et al (2006) were used: a) Intra sentential pause number and duration, b) Inter sentential pause duration, c) f0 decline at the EOS. Both Group A and Group B recordings were measured for the above features using PRAAT. 10 Adult fluent readers of Indian English were also recorded reading out the same passages and these recordings were also analysed as a baseline measure.

There was a large variation in the results for each student and unlike Mostow & Duong, (2009) we found that individual student results were not statistically significant. However, when clubbed together as Group A and Group B, the average scores across the two groups showed significant improvement for Group A for inter and intra sentential pause measures. Table 1 (a) and (b) show the results for Intrasentential pauses for both the frequency of occurrences and duration for both the groups and the adult averages as a baseline. While both groups have a decrease in both the values, a two-tailed t-test showed that in the case of Group B these are not statistically significant. However, it may be noted that the numbers are significantly higher than the baseline adult measurements.

	Avg length of pauses before RT	Avg length of pauses after RT	
Group A	546 ms	402 ms	$p<0.05$
Group B	758 ms	743 ms	$p=0.9$
Adult	28 ms		

	Avg no. of pauses before RT	Avg length of pauses after RT	Statistical significance
Group A	12	8	$p<0.05$
Group B	15	11	$p=0.78$
Adults	3		

Table 1 (b): *The average length of Intrasentential pauses*

The pause lengths for both inter and intra sentential pause lengths were marked by visual observation of a spectrogram and the corresponding waveforms. The duration of each pause was noted in milliseconds and averaged across each sentence.

In the case of intersentential pauses, we observed that many students did not pause at all between consecutive sentences especially in the pre-test recordings. Post-test recordings showed that many students in Group A started pausing in between sentences. This is contrary to the results in Schwanenflugel et al (2006) that show that an increase in fluency and decoding is strongly correlated with shorter pauses in between sentences and more reliance on f0 declination. However, in the participants of this study the reading skills were so low that they often did not recognize end of sentences and hence failed to mark them. With increase in skill levels, the correct decoding of an EOS caused many to demarcate it with a short pause. The results for Intersentential measurements are presented in Table 2. Again, though both groups show the same general trend, the results for Group A are statistically significant.

	Avg pause length before RT	Avg pause length after RT	
Group A	186 ms	329 ms	$p<0.05$
Group B	176 ms	184 ms	$p=0.82$
Adult	18 ms		

Table 2: *The average length of intersentential pauses*

We also measured the fall in f0 at the end of sentences as a measure of more fluent reading. This declination was measured as a fall from the previous peak in the sentence to the end of the sentences. Both Group A and Group B showed very flat f0 contours at EOS, and there was no significant change in this number pre or post-test. Again this was at odds with the adult baseline measure. An interesting observation on Group B recording was a significant number of high f0 at the end of the sentence. This was mainly due to the presence of a continuation contour at the end of the sentence. Table 3 shows the results for f0 declination at the end of the sentence for both groups and the baseline adult measurement. Adults were also not consistent in the no. of pauses but most showed a declination pattern for end of sentence structure.

	F0 fall at EOS before RT	F0 fall at EOS after RT	
Group A	-14	7	p=0.57
Group B	-21	-19	p=0.89
Adult	75		

Table 3: *The f0 declination at the end of sentence*

We tried to measure the correlation between each of the above measures with CBM scores at the individual level. However, no significant correlation was found with CBM. This may be due to the coarse granularity of CBM itself as a measure or because of the great variability in the data across the students. The above results show that while there is no significant individual improvement in prosodic reading, if clubbed together then the students in Group A show a lower no. of intra-sentential pauses as well as shorter intra and inter sentential pauses than Group B. However, both the groups show a flat f0 at the end of declarative sentences. The rising (continuation) f0 at the EOS for Group B coupled with shorter pauses indicate that the students in Group B are not processing the end of sentence at all. Also, Group B students showed no significant change in their prosodic reading even after consistent use of the reading tutor clearly while there is a statistically significant positive gain for Group A student help us conclude that there is a minimal level of decoding skill required before any positive changes in reading prosody can kick in.

As far as prosodic chunking is concerned, the intrasentential level were too numerous and lengthy to support any clear evidence for attempt at chunking the words into meaningful phrases. Most readers in Group B were reading in a hesitant, word-by-word manner though there were some students in Group A whose reading showed some instances of a more natural grouping of words.

Finally, we found no evidence in the data from Group A that their reading prosody was closer to that of the adults. Infact, a large variance was found between the adult reading data and the data from both the groups. Duong & Mostow (2009) also found that correlation with adult data was not a very good indicator of improvements in reading prosody. This may be because even though Group A was more fluent as reflected in their reading prosody as well as CBM results, their developmental stage was still far from that of a fluent ESL reader.

However, the results do provide support for the hypothesis that the use of the reading tutor does significantly improve the reading prosody of the students, at least for Group A students. It may be noted that Group A were also students who showed a higher and more significant changes in their decoding skills as tested by the CBM evaluation. The disparity in the results between Group A and Group B are not surprising but reflect the fact that prosodic fluency is an effect of decoding fluency. That is, more expressive and appropriate use of prosody in reading comes as decoding accuracies get better.

Conclusion

In this paper, we have looked at three prosodic features of reading prosody for L2 (ESL) learners at elementary school level, namely, inter and intrasentential pauses, and f0 declination at EOS, to determine improvements in prosody through the use of speech recognition based reading tutor. The results show that there are definite prosodic fluency gains with the use of Reading tutor. However, these gains are a function of the gains in decoding skills. The students in our study could clearly be divided into two groups based on their decoding skills and the group with higher

improvements in decoding accuracies also showed a significant improvement in their prosodic reading skills. As pointed out by Sitaram and Mostow (2012), we also found that comparison with adult fluency data is not a good feature for our set of students because their skills are still at such low developmental levels for any meaningful comparison.

This is a preliminary investigation and any future work would have to look at more prosodic cues. One shortcoming of the study was the coarseness of CBM as a metric for decoding skills. We could find no correlation between the CBM scores of individual students and the prosodic features. A more granular score of reading fluency that takes into account hesitations, self-corrections, mispronunciations, content versus functional word decoding, etc. as different features could perhaps offer a better correlation with the prosodic features.

As far as we know, this is the first study of prosodic features for ESL in Indian English and we have not looked into any interference from mother-tongue or L1 prosody in this data. This is primarily due to the very low levels of reading skills for these students. It would be interesting to see if readers with higher fluency exhibit chunking and/or f0 contour patterns which show influences of their mother-tongue.

Acknowledgements

The author would like to thank the students and the staff of the three participating schools for their time and effort for conducting the study.

References

Agnihotri, R., and Khanna, A., English Language Teaching in India; Issues and Innovations, Sage Publications, 1995.

Boersma, P. and D. Weenink. Praat: doing phonetics by computer (Version 5.0.33) [Computer program downloaded from http://www.praat.org/]. 2008.

CMU Sphinx: http://cmusphinx.sourceforge.net/

Cruttenden A. Intonation comprehension in 10-year-olds. Journal of Child Language. 1985;12:643–661.

Cutler A, Swinney DA. Prosody and the development of comprehension. Journal of Child Language. 1987;14:145–167.

Deno, S. L. Curriculum-based measurement: The emerging alternative in Exceptional Children, 52(3):219–232, 1985.

Duong, M., & Mostow, J. (2009, September 3-5). Detecting Prosody Improvement in Oral Rereading. Second ISCA Workshop on Speech and Language Technology in Education (SLaTE), Wroxall Abbey Estate, Warwickshire, England.

Kuhn, M. R., & Stahl, S. A. (2004). Fluency: A review of developmental and remedial practices. In R. B. Ruddell & N. J. Unrau (Eds.), Theoretical models and processes of reading (5th ed.) (pp. 412–453). Newark, DE: International Reading Association.

Kuhn, M. R., & Schwanenflugel, P. J. (Eds.). (2008). Fluency in the classroom. New York: The Guilford Press.

Larsen, S., Hammill, D., and Moats, L. Test of Written Spelling. Fourth Edition, Pro-Ed, Austin, Texas, 2009.

Mostow, J. and G. Aist. Evaluating tutors that listen: An overview of Project LISTEN. In K. Forbus and P. Feltovich, Editors, Smart Machines in Education, 169-234. MIT/AAAI Press: Menlo Park, CA, 2001.Schwanenflugel, P. J., & Ruston, H. P. (2008). Becoming a fluent reader: From theory to practice. In M. R. Kuhn & P. J.

Mostow, J., & Duong, M. (2009, July 6-10). *Automated Assessment of Oral Reading Prosody*. Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED2009), Brighton, UK, 189-196.

Palumbo, T. J., & Willcutt, J. R. (2006). Perspectives on fluency: English-language learners and students with dyslexia. In S.J. Samuels & E. Farstrup (Eds.), *What research has to say about fluency instruction* (pp. 159–178). Newark, DE: International Reading Association.

Ramanathan, H., Testing of English in India: A developing concept, in Language Testing, 25(1). 2008. 111-126.

Schwanenflugel, P.J., E.B. Meisinger, J.M. Wisenbaker, M.R. Kuhn, G.P. Strauss, and R.D. Morris. Becoming a fluent and automatic reader in the early elementary school years. *Reading Research Quarterly*, 2006. *41*(4): p. 496-522.

Schwanenflugel (Eds.), Fluency in the classroom (pp. 1–16). New York: The Guilford Press. *2007.*

Sheorey, Ravi, Learning and Teaching English in India, Sage Publications, 2006

Sitaram, S., and Mostow, J., Mining Data from Project LISTEN's Reading Tutor to Analyze Development of Children's Oral Reading Prosody. FLAIRS 2012, Florida.

Weber, F and Bali, K. Enhancing ESL education in India with a reading tutor that listens. In Proceedings of the First ACM Symposium on

Computing for Development (ACM DEV '10). ACM, New York, NY, USA, 2010.
Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. Scientific Studies of Reading, 5, 211–239.
Yamashita, J & Ichikawa, S. "Examining reading fluency in a foreign language: Effects of text segmentation on L2 readers". In Reading in a Foreign Language, Volume 22, No. 2, pp. 263–283, Oct 2010