Reading Companion: The Technical and Social Design of an Automated Reading Tutor

Arthur Kantor[1], *Miloš Cerňak*[1†], *Jiří Havelka*[1],
Sean Huber[2‡], *Jan Kleindienst*[1], *Doris B. Gonzalez*[3]

[1]IBM Research Lab, Prague, Czech Republic
[2]IBM Global Business Services, Chicago, IL, USA
[3]IBM Corporate Citizenship & Corporate Affairs, Armonk, NY, USA

Abstract

This paper describes IBM's automatic reading tutor system, the Reading Companion. The reading tutor aims to improve the literacy skills of beginning readers, both children and adults, and help adults who are non-native speakers of English to learn the language. We describe Reading Companion's architecture, which allows a large, globally distributed reading companion community to create and share new reading material. We also report substantial accuracy improvements in recognizing children's speech gained by training the recognizer on the IBM Kidspeak corpus, a newly developed corpus of children's speech.

Index Terms: computer aided learning, child speech recognition, automatic reading tutor

1. Introduction

The ability to read is essential for people to live productive and rewarding lives. In developed and developing countries, many kids and adults read below levels considered proficient for their age. For example in the United States, 33% of fourth-graders read at "below basic" achievement level [1].

To help address this problem, IBM started Reading Companion [2] as a charity project. Reading Companion (RC) is a web-based software that uses speech recognition technology to help children and adults learn how to read. The software is available for free to public elementary schools (for children ages five through seven) and nonprofit organizations such as public libraries, community colleges, and agencies that offer adult literacy services. The software is available to sites worldwide, with more than 126,000 registered users from 2,778 sites (half of which are schools) from 40 countries taking part in the grant program. The project is ongoing and is now sixteen years old.

The goals and the user interface of Reading Companion is similar to other existing projects such as CMU's Project Listen [3], Colorado Literacy Tutor [4], and some others [5, 6]. A variety of carefully designed studies show that using these reading tutors increases the learning gains over the gains children would make otherwise. For example, on the task of vocabulary learning, Project LISTEN's tutor improved the test score more than regular classroom activities and, surprisingly, was as effective as one-on-one tutoring by certified teachers [7].

Gains in child reading ability were not measured for RC. However RC was compared against typical classroom instruction of English-as-a-second-language (ESL) adults, where the students were taught basic job interview skills, and then evaluated with a test developed by the welfare-to-work program educators. There was no statistical difference between RC instruction and human teacher instruction, and both groups of students significantly outperformed the group that received no instruction at all [8]. While no piece of software can replace a teacher, the above studies show that RC and similar automatic tutors can be effective in improving certain reading skills of both children and adults.

In order to maximize the benefit of automatic tutors to the world-wide literacy effort, it is important to make the software accessible and relevant to the emerging readers worldwide. For the tutor to be accessible, it needs to be available on a variety of platforms, especially mobile and low-cost devices as well as devices with low-speed internet connections. The software also needs to be upgradeable with minimum effort from the users, since many teachers and students do not have advanced computer skills.

It is also important to keep the software relevant, so the students remain interested in using the tutor. This can be done by dynamically adjusting the task difficulty level, by making the tutor appear game-like, by providing in-game encouragement and awards, and by continually presenting new age- and culture-appropriate reading material as well as material related to other classroom activities. Allowing the students and teachers to create and share new reading material with other users is an effective and low-cost way to keep the tutor software relevant to the students and teachers alike.

This paper describes the social context within which RC project exists and the technical design choices made to make RC accessible and relevant to emerging readers. Our experience with RC may be relevant to other automated tutor projects which try to maximize their impact on world literacy.

This paper makes three novel contributions.

- We describe a long running (16-year) global program to improve reading literacy, and the various educational scenarios in which RC was found useful. We also discuss how we enabled the RC community (educators and students) to create and share new reading material with other RC users.
- We describe the RC infrastructure which allows us to make RC available on a variety of platforms, including mobile devices, and allows us to collect speech as a byproduct of normal RC use, which can be used to further improve the RC recognizer.
- We use a newly developed ∼80 hour kids' speech corpus collected from ∼800 students to substantially improve

[†]Presently at Idiap Research Institute, Martigny, Switzerland.
[‡]Presently at Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA.

the accuracy of our speech recognizer compared to the recognizer trained on adult speech and adapted to children's speech.

The remainder of this paper is organized as follows. In Section 2, we describe the social goals of Reading Companion for students, teachers and parents. Section 3 describes RC from the student's point of view: what kind of immediate feedback (encouragement, assistance and corrections) is provided by RC to the student, how the student speech is evaluated, how the level of longer-term assistance (scaffolding) changes in response to the perceived student reading performance and which aspects of student speech and behavior drive the behavior of the RC. In Section 4, we discuss the the global RC program: the use cases, the features popular with teachers, content authoring and infrastructure choices to make RC as widely available as possible. Section 5 describes the characteristics and configuration of the automatic speech recognition (ASR) engine. In Section 5.1, we also describe the new IBM Kidspeak corpus as well as the accuracy of the recognizer trained on the new corpus. Section 6 discusses the future work and concludes.

2. Social Design Goals of Reading Companion

An automated tutor can affect the behavior of students, teachers and to a lesser extent, the students' parents. In this section, we discuss the social design goals we consider important for an automated tutor, while leaving to the following sections the details of how these goals influenced RC's implementation.

2.1. Students

For students, automated tutors provide an opportunity to reward reading practice and individual improvement instead of rewarding the achievement of some set reading level. They can do this through self-paced, private and individualized instruction.

With self-paced instruction, students can repeat exercises as often as necessary, without the feeling of wasting the teacher's time.

Since using a computer or a tablet is a semi-private activity, students should feel less performance anxiety than reading publicly in a typical classroom setting. This is particularly important for under-performing or adult readers who may feel ashamed or inadequate when reading in a group.

A well designed automated tutor should also keep the students engaged and motivated. This can be done through individualized instruction, where the tutor is designed to focus on the pronunciation and comprehension difficulties specific to the student and adapt the difficulty to the ability of the student.

Virtual awards or badges can also be powerful motivators [9]. Most directly, badges can be used to reward desired behavior, e.g. awarding progressively more rare and valuable badges for progressively longer time spent reading. Indirectly, badges also provide a sense of group identity among the students.

If the badges are made public, they can also be used to communicate status and reputation to fellow students and promote competition on the desired behavior (reading practice) rather than on absolute reading level. The value of competition among students probably depends on the circumstances and the decision of whether or not to make the badges public can be left to the teacher.

To summarize, a well designed automatic tutor can improve a student's attitude towards learning to read by removing the stigma of making public mistakes, by adapting the difficulty to keep the student challenged and by rewarding practice and individual improvement rather than achievement of absolute goals.

2.2. Teachers

Teachers often express interest in having reading material that is appropriate to their classroom. An automated tutor should make it easy to create, disseminate and recommend new reading material. Ideally, the teachers would be allowed to collaboratively author the books. For example, many of the teachers using Reading Companion are not native speakers so they may ask some other native-speaking teacher to record the prompts for the book.

2.3. Parents

The parents' involvement in the child's schooling is a major factor influencing the success of the education. As mobile devices become more pervasive, it becomes reasonable to think of using an automated tutor as an activity that can take place at home as well as in the school. Using an automated tutor can be a common activity to be shared by student, teacher and parent, thus bringing the parents more into the child's education process.

Finally, an automated tutor can make an excellent research platform within which teaching techniques and motivators can be rigorously evaluated. The ability to perform educational experiments is another worthy goal for an automated tutor.

In the next sections, we present the design of Reading Companion which tries to reach the social goals we have just described.

3. The Student's View of Reading Companion

In this section we describe the student's interaction with RC. We describe the user interface, the way we evaluate the student's reading performance, and the way we change the type of assistance and feedback in response to the student's evolving reading skills.

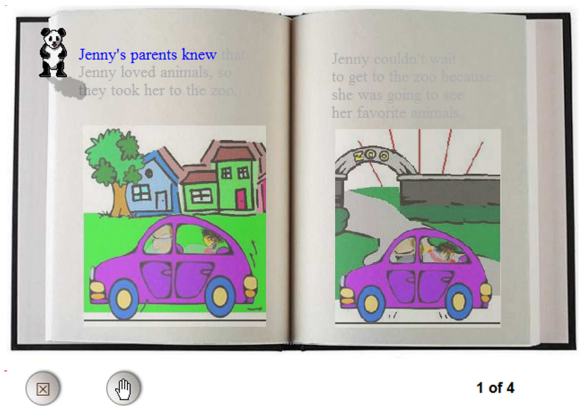


Figure 1: Reading Companion from the student's point of view.

Figure 1 shows a screenshot of the student interface and a more detailed discussion of user interface considerations can be found in [10]. The student is guided by an animated character (a panda for children, a homunculus for adults), who walks around the book to draw attention to the phrases the student will work with. Depending on how the book was authored, the guide may then speak a short introduction to the page. Depending on the current level of assistance (scaffolding level), the guide may also read the whole page to the student.

At this point the reading begins, one phrase at a time. Each phrase is highlighted in blue, and depending on the scaffolding, may be partially read by the guide. The guide then prompts the student to read by putting its paw or hand to its ear and highlighting in red the words to be read. The end of speech is automatically detected by the recognizer, but can also be indicated by the student clicking on the guide. Listening also stops if no speech is detected for more than 8 seconds.

If the phrase is read correctly, the guide gives a brief encouragement (e.g. "Great!" or "That's it!") and continues with the next phrase. Otherwise, the guide highlights the wrongly pronounced words one at a time and prompts for a repeat reading attempt. If the second attempt on the individual words is correct, a brief encouragement is spoken. Otherwise, the guide models (speaks) the problematic word and moves on to the next phrase. This approach masks ASR errors by never explicitly giving negative feedback to the student.

At any time, the reading session can be ended (by clicking on the close button), and will be resumed from the same spot once the book is opened later. The guided reading can also be paused. In the paused mode, the student can click on any word to have it spoken by the guide.

3.1. Reading Evaluation

We now describe how RC decides whether a prompted word is spoken correctly or incorrectly.

The decision of which words in the prompt phrase are marked incorrect depends on both the recognition grammar and the comparison of the recognized phrase to the reference phrase. The grammar is a simple $word*$ grammar which allows the recognizer to generate hypotheses with the regular expression $word_1* \;\; word_2* \; \ldots \; word_{\mathrm{n}}*$ for the reference phrase $word_1 \;\; word_2 \; \ldots \; word_{\mathrm{n}}$. The same grammar is used even if the reader is prompted for only a part of the phrase due to scaffolding. Phoneme loops are used as a garbage pronunciation and background model.

A reference word is marked correct simply if it is present in the hypothesis. If we assume that the words in the reference phrase are unique and the recognizer is accurate, then a subsequence of the reference phrase will be marked correct exactly when it is a subsequence of the hypothesized phrase. This means that deletions are counted as errors, insertions are not, and substitutions will be counted as errors only if they are acoustically closer to the preceding or following word than to the target word.

3.2. Scaffolding

RC gives more assistance (scaffolding) to readers when it detects that they are having difficulty reading the phrases. As readers become more proficient, the scaffolding is removed. RC defines four reading proficiency levels:

1. All words in a phrase are modeled. The student repeats the last two words in a phrase.
2. All words except the last two in a phrase are modeled. The student reads the last two words in a phrase.
3. All words in a phrase are modeled. The student repeats all the words in a phrase.
4. No words are modeled. The student reads all the words in a phrase.

By default, a student's reading level automatically adjusts (starting from level two) based on how well the student performs. As the student reads more words correctly, the reading level increases so the student practices reading larger portions of the text. If a student makes too many reading errors, the reading level decreases so more words are modeled and possibly the student reads shorter portions of text.

Adaptive scaffolding, along with a large collection of books, makes RC relevant to students over a broader range of reading abilities as the students become proficient readers.

4. The Reading Companion Program

RC was originally designed to improve the reading skills of children, primarily kindergarten through 3rd grade students, who come to school with less pre-literacy experience, or have some difficulty participating in whole and small group reading activities. Since then, RC has also found use in adult literacy programs and in teaching English to students in non-English speaking countries. In Section 4.1 we describe RC from the teacher's point of view, as well as the book authoring process through which the teachers can create and share new RC books. In the following section, we also talk about which RC features the teachers found useful (or not). In Section 4.3 we describe the infrastructure which is designed to make RC as accessible as possible.

4.1. The Teacher's View of Reading Companion

The RC access is controlled hierarchically, with RC access permissions passing from the Reading Companion administrator to the responsible person at the grant site (such as a school) to the actual teachers.

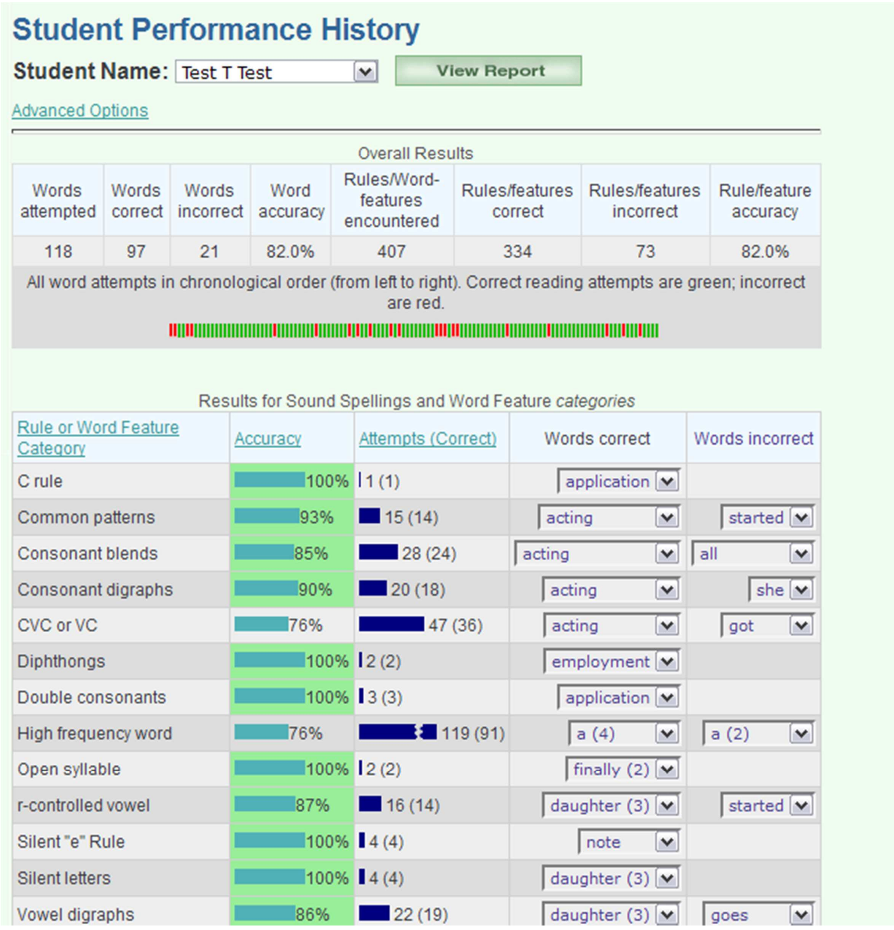
The teachers can create classrooms and populate them with students. They also specify whether the classroom is for adults or children, and this decides which acoustic model will be used by the RC speech recognizer and also what kind of reading material is available for the classroom. The teachers can also fill the classroom bookshelf with interactive books from the global RC library and these books are made available to the students once they log in. Reading companion collects performance statistics from each student, and the teachers can generate reports from these statistics.

The student performance reports group the student mistakes by word type, and also by *word feature category*. For example, the 'c rule' word feature category would include all the words where the letter 'c' is pronounced as the phoneme `/K/`. Figure 2 shows an example of a detailed performance report. The word feature categories are all derived from lexical analysis only, and do not depend on the speech recognizer.

The analysis of mistakes grouped by word feature category may be useful in automatic selection of future student exercises, but many teachers and instructors found the student evaluation page intimidating and indicated that they wanted to see more information on how their class was progressing as a whole and wanted to receive the information in a more simplified way. RC now generates these simplified class-wide reports as well.

4.2. Authoring

Another feature popular with teachers is the ability to author their own books which are relevant to their class. RC contains a flexible authoring tool called Book Builder which allows teachers to write their own books with images and text, and also to record the audio which will be played back when the student



Student Performance History

Student Name: Test T Test View Report

Advanced Options

Overall Results

Words attempted	Words correct	Words incorrect	Word accuracy	Rules/Word-features encountered	Rules/features correct	Rules/features incorrect	Rule/feature accuracy
118	97	21	82.0%	407	334	73	82.0%

All word attempts in chronological order (from left to right). Correct reading attempts are green; incorrect are red.

Results for Sound Spellings and Word Feature *categories*

Rule or Word Feature Category	Accuracy	Attempts (Correct)	Words correct	Words incorrect
C rule	100%	1 (1)	application	
Common patterns	93%	15 (14)	acting	started
Consonant blends	85%	28 (24)	acting	all
Consonant digraphs	90%	20 (18)	acting	she
CVC or VC	76%	47 (36)	acting	got
Diphthongs	100%	2 (2)	employment	
Double consonants	100%	3 (3)	application	
High frequency word	76%	119 (91)	a (4)	a (2)
Open syllable	100%	2 (2)	finally (2)	
r-controlled vowel	87%	16 (14)	daughter (3)	started
Silent "e" Rule	100%	4 (4)	note	
Silent letters	100%	4 (4)	daughter (3)	
Vowel digraphs	86%	22 (19)	daughter (3)	goes

Figure 2: Student performance reports as viewed by the teacher. The history of word attempts, successes and failures is in the top table. Attempts grouped by word feature category are in the rows of the second table. For each word feature category, the accuracy and the attempted word types belonging to the category can be seen.

needs assistance. The book in Figure 1 is an example of teacher generated content.

The written text can be segmented into phrases (by inserting a | character at phrase boundaries), and the teacher is asked to record audio for each phrase, and also for each word individually. The guide's spoken instruction prompts need to be recorded separately for each book as well. The result is a consistently voiced, natural sounding speech which does not unnecessarily distract the student throughout the reading session.

Once the author is satisfied with the book, she describes it with some metadata: a target age-level, topic category, reading level and geographical region, and submits it to the RC book committee for review and publication in the RC library. The teachers are able to rate the quality of the books, and search for books by quality and metadata. Many authors are non-native English speakers themselves, and they sometimes ask the book committee to record the audio for their books.

Once the book is published, the whole RC community is granted access to it. Currently, the library contains about 400 books, and is growing at a rate of 30 books per month. It contains book collections for children, collections on financial literacy, housing, job search, court probation, U.S. citizenship, driving and also narrative and vocabulary collections.

4.3. Infrastructure

We want RC to be easy to install and upgrade and to be available on a variety of types of devices including portable devices with low computational resources. We also would like to make RC accessible from school and from home. This suggests that the user interface should be portable and lightweight, where the audio is sent to a remote ASR server for computationally intensive speech recognition. For most situations, this is the approach we take.

Figure 3 shows the RC's distributed architecture. The administrative and teacher user interface is implemented as a typical web application, while the student user interface is implemented in Adobe Flash, so it can run within all popular web browsers and operating systems as well as Android and Apple smart phones. The client receives the book material, including all the audio prompts and recognition grammars from the web server. At the end of a reading session, the student performance information and a bookmark is sent to the server. The bookmark allows the student to resume the book at the same place with the same scaffolding level during the next reading session.

Before RC prompts the student to read a phrase, it sends a phrase-specific recognition grammar to the ASR server. During listening, the audio is captured by Flash, optionally compressed

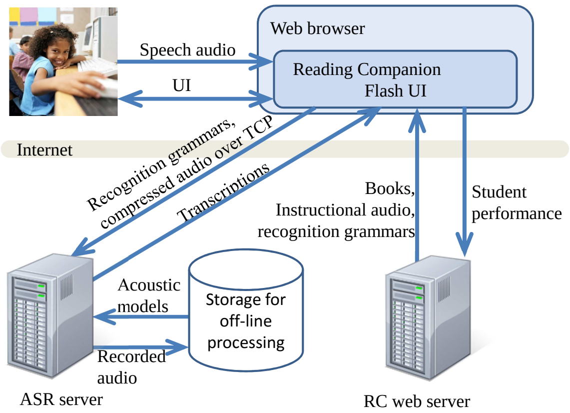


Figure 3: Infrastructure for the Reading Companion.

using the Speex codec and is streamed to the ASR server. Depending on the level of compression, the required bandwidth is up to a maximum of 5.2 kBps for each speaking student.

The recognition can be stopped either by the student or by the ASR server if it detects the end of speech, at which point the transcription and speech quality information can be sent back to the client. The audio can be saved at the server for off-line retraining of the recognition models.

To minimize student distractions, we need the user interface to stay responsive. In telephony applications, audio latency above 300ms is considered unacceptable. If we adopt the same standard, once the end of speech is detected, total round-trip internet and recognition latency must fall under this limit. Such low latency response is possible, but it requires a high speed internet connection which many of the schools in developing countries do not have. Some of our earlier experiments with server-based ASR revealed latencies in schools that varied from 1 to 32 seconds [10].

For these distant sites, we offer an RC installer which includes a speech recognizer, acoustic models, and browser plug-ins, so that the recognition takes place on the client machine. The unavoidable disadvantage of this configuration is that some computer skills are required to install the recognizer, the student speech cannot be captured and the installer package is available for a smaller range of devices (currently only for Microsoft Windows and Linux). As the internet connectivity improves worldwide, we expect that the majority of the sites will be able to use the remote ASR servers.

5. Automatic Speech Recognition in Reading Companion

This section describes the RC speech recognizer. The IBM Kidspeak corpus of children's speech which was used to train the child-specific acoustic models is described first. The recognizer itself and the evaluation of the acoustic models is described in Sections 5.2 and 5.3.

5.1. The IBM Kidspeak Corpus

The IBM Kidspeak Corpus is a corpus of American English children's speech, collected between 2006 – 2009 at schools belonging to the Chicago and New York areas. Altogether 799 children from 6 to 9 years old participated in the recording sessions (6.3 minutes of speech per child on average).

The speech was recorded at 16bits/sample 22.05kHz with a headset microphone. About half of the recordings were done in a natural classroom setting with some background noise (but without overlapping speech from the instructor / recording technician). The remaining recordings were made in a quiet room. The recording was done with a modified version of RC itself, with the speech prompts selected from existing RC books.

The training and testing data were collected using two protocols: the 'clean' protocol with the goal to record correct readings, and the 'error' protocol with the goal to record correct and incorrect readings of the prompts at natural frequencies.

- **The 'clean' protocol.** The child reads the text on the screen. If it is correct, that recording is used. Otherwise the instructor tells the child how to say it correctly, and another recording is attempted. If there is no correct reading after 2 or 3 attempts, that prompt is skipped and the child moves on to the next prompt.
- **The 'error' protocol.** The child reads the text on the screen once, possibly with mistakes, and the first attempt is used. Roughly half the utterances had some sort of error.

There was no speaker overlap between the two sets of recordings. Table 1 shows the statistics for each protocol.

Table 1: *The number of speakers, number of utterances, hours of speech, number of unique words and number of phrases recorded with the 'clean' and 'error' protocols for the IBM Kidspeak corpus.*

Protocol	Spkrs	Utts	Hours	Words	Phrases
Clean	636	65642	66.5	8487	1977
Error	163	14425	20.8	4911	1106

The texts were chosen from the RC books available at the time, with the main emphasis on phonetic representativeness and variety. We also tried to minimize the overlap between the sets of books used for the 'error' and 'clean' protocol recordings. The objective function that we minimized was based on the Bhattacharyya distance of phone unigram and phone bigram distributions in the selected texts to the respective distributions in a large text corpus (we used the Open American National Corpus [11]) and on phone bigram coverage (phone unigram coverage came for free).

For the entire corpus, the prompts can be used as word-level transcriptions. Additionally, for the error utterances, a phonetic transcription was automatically generated from the single most common pronunciation from the dictionary. A human transcriber without any specialized linguistic background then listened to each utterance and for each miscue in the utterance corrected the canonical phonetic transcription so it reflected the actual speech.

5.2. ASR

RC uses the IBM Embedded ViaVoice (EVV) speech recognition engine [12] for real-time decoding of speech. EVV is designed for grammar-based command and control applications with medium to large vocabularies. A sentence-specific weighted grammar is created for each phrase and is combined with a garbage model. A minimized and determinized finite state graph is compiled on the fly using a grammar compiler, and is then used for efficient decoding.

The decoder can detect the end of speech automatically by checking if the finite state machine spent sufficient time in the same accepting state of the decoding graph. Some additional heuristics are used to make the end of speech detection more accurate.

EVV can perform one-pass decoding and also decoding with unsupervised fMLLR [13] adaptation. We found fMLLR especially useful for the environment adaptation (see Section 5.3), since the children were recorded in two conditions (quiet and classroom environments).

The speech recognizer typically uses a pronunciation dictionary, but if a word is missing in the dictionary, it can generate a phonetic pronunciation given the word's spelling, similar to the approach described in [14]. This feature is especially useful if new e-books are added into the RC library and the pronunciation dictionary of the recognizer is not updated.

5.3. ASR Evaluation

The adult speech recognition uses a speaker independent acoustic model with a size of 60k Gaussians, discriminatively trained with the minimum phone error (MPE) objective. To recognize children's speech, RC for years had used the adult acoustic models adapted to children's speech. This is the baseline in our evaluation (see Table 2). We have been getting complaints from the teachers about the RC's accuracy, and the need for improvement was apparent. In 2010, we developed HMM/GMM systems using IBM Kidspeak corpus, and now report their performance compared with the baseline.

All evaluated systems used three-state, word-internal triphone models, trained from 39 dimensional MFCCs (12 cepstral plus energy coefficients) including delta and delta-delta features. Training of new acoustic models was performed with IBM's Attila speech recognition toolkit [15] on the 'clean' training set of the Kidspeak corpus. Both baseline and new acoustic models were compared on the 'error' testing set of the Kidspeak corpus, simulating the RC recognition: grammars in the test were built dynamically for each sentence in the same way as they are built in the reading tutor.

Table 2: *Word error rates (WER) and Sentence error rates (SER) of children acoustic modeling on data with reading miscues.*

Model	WER	SER
Baseline	25.8%	46.7%
ML 60k	12.5%	28.9%
MPE	10.0%	25.9%
BMMI	10.0%	25.7%
BMMI fMLLR	9.3%	23.7%

Table 2 shows recognition results achieved on models trained with progressively more complex algorithms. We started with children's Maximum Likelihood (ML) models initialized from the existing baseline model, achieving significant improvement. While the older baseline recognition system used as features MFCC and their deltas, the new speech recognition system used LDA features calculated from the current, four previous and four following frames. In addition, we also successfully decreased the model size by almost half by the application of the Bayesian Information Criterion (BIC). The change in the algorithms alone is not enough to account for the 50% relative WER reduction, and the remaining WER difference must come from using a training corpus better matched to the testing conditions.

ML training was followed by discriminative training. Boosted MMI (BMMI) [16] was found to be the best discriminative method. The best performance was observed on BMMI models with decoding with fMLLR adaptation, which mainly compensate for environment variations.

6. Conclusion

In this paper, we described the technical and social issues encountered while developing Reading Companion, an automatic reading tutor with a large, globally distributed user base. Our main goals were to make the system as accessible and relevant as possible.

To that end, Reading Companion is designed to run on as many devices as possible, with minimum installation, while also being able to run in sites with poor internet connectivity. Reading Companion tries to be relevant to a wide range of emerging readers by dynamically adapting to the students' reading abilities, and by allowing teachers to create new reading material that is appropriate for different age and culture groups.

While Reading Companion's user interface is designed to mask the imperfect accuracy of the existing speech recognition systems, we continue to work on improving the accuracy of our recognizer. We developed the IBM Kidspeak corpus, an 80-hour corpus of kids' speech which we then used to substantially improve the accuracy of our acoustic models.

6.1. Future Work

The development of Reading Companion is continuing. Besides continuing to improve recognition accuracy, we are planning to extend RC's capabilities to pronunciation tutoring.

Syllable stress plays an important role in efficient spoken communication in English, as the meaning of a word can change based on its stress pattern (e.g., *address* or *content*). Many foreign languages lack intra-word syllable stress. This results in many foreign speakers of English either not stressing any of the syllables of a word or stressing the wrong syllable. For the benefit of ESL students, we plan to extend RC to provide syllable stress instruction by using a syllable-stress evaluation algorithm such as the one described in [17].

7. Acknowledgments

We would like to thank Petr Novák, who helped us with building the child acoustic models, and the many colleagues who contributed or contribute to the development of Reading Companion.

8. References

[1] National Center for Education Statistics, "The nation's report card - reading 2011," http://nationsreportcard.gov/reading_2011/, 2011, [Online; accessed 14-June-2012].

[2] IBM, "The Reading Companion website," http://www.readingcompanion.com/.

[3] J. Mostow, "Why and how our automated reading tutor listens," in *International Symposium on Automatic Detection of Errors in Pronunciation Training (ISADEPT)*, June 2012, pp. 43–52.

[4] Barbara Wise, Ron Cole, Sarel Van Vuuren, Scott Schwartz, Lynn Snyder, Jariya Tuantranont, and Bryan Pellom, "Learning to read

with a virtual tutor: Foundations to literacy," in *Interactive Literacy Education: Facilitating Literacy Environments through Technology.*, C. Kinzer and L. Verhoeven, Eds. 2005.

[5] Marilyn J. Adams, "The Promise of Automatic Speech Recognition for Fostering Literacy Growth in Children and Adults," in *International Handbook of Literacy and Technology*, Michael C. McKenna, Ed., vol. 2, pp. 109–128. Routledge, Apr. 2006.

[6] Don Nix, Peter Fairweather, and Bill Adams, "Speech recognition, children, and reading," in *CHI 98 conference summary on Human factors in computing systems*, New York, NY, USA, 1998, CHI '98, pp. 245–246, ACM.

[7] Gregory Aist, Jack Mostow, Brian Tobin, Paul Burkhead, Andrew Cuneo, Brian Junker, and Mary Beth Sklar, "Computer-assisted oral reading helps third graders learn vocabulary better than a classroom control - about as well as one-on-one human-assisted oral reading.," in *Proceedings of the Tenth Artificial Intelligence in Education (AI-ED) Conference*, 2001.

[8] Paula M. Bach and Jennifer Lai, "Usability and learning in a speech-enabled reading tutor: a field study," in *CHI '06*, New York, NY, USA, 2006, pp. 502–507, ACM.

[9] J. Antin and E.F. Churchill, "Badges in social media: A social psychological perspective," in *CHI 2011, ACM, Vancouver, BC*, 2011.

[10] Keith Grueneberg, Amy Katriel, Jennifer Lai, and Jing Feng, "Reading Companion: A interactive web-based tutor for increasing literacy skills," in *Human-Computer Interaction – INTERACT*, 2007, pp. 345–348.

[11] Randi Reppen, Nancy Ide, and Keith Suderman, "American national corpus (ANC) second release," `http://americannationalcorpus.org/OANC/`, 2005.

[12] Tomáš Beran, Vladimír Bergl, Radek Hampl, Pavel Krbec, Jan Šedivý, Bořivoj Tydlitát, and Josef Vopička, "Embedded ViaVoice," in *TSD*, Petr Sojka, Ivan Kopecek, and Karel Pala, Eds., vol. 3206 of *LNCS*, pp. 269–274. Springer, 2004.

[13] Sreeram V. Balakrishnan, "Fast incremental adaptation using maximum likelihood regression and stochastic gradient descent," in *EUROSPEECH*, 2003, pp. 1521–1524.

[14] Benoit Maison, "Automatic Baseform Generation from Acoustic Data," in *EUROSPEECH*, 2003, pp. 2545–2548.

[15] Hagen Soltau, George Saon, and Brian Kingsbury, "The IBM Attila speech recognition toolkit," in *IEEE Workshop on Spoken Language Technology*, Dec. 2010, pp. 97–102.

[16] Daniel Povey, Dimitri Kanevsky, Brian Kingsbury, Bhuvana Ramabhadran, George Saon, and Karthik Visweswariah, "Boosted MMI for model and feature-space discriminative training," in *ICASSP*, 2008, pp. 4057–4060.

[17] Harish Doddala, Om D. Deshmukh, and Ashish Verma, "Role of nucleus based context in word-independent syllable stress classification," in *ICASSP*, 2011, pp. 5712–5715.