

Giving Help and Praise in a Reading Tutor with Imperfect Listening –
Because Automated Speech Recognition Means Never Being Able to Say You're Certain

Abstract

Human tutors make use of a wide range of input and output modalities, such as speech, vision, gaze, and gesture. Computer tutors are typically limited to keyboard and mouse input. Project LISTEN's Reading Tutor uses speech recognition technology to listen to children read aloud and help them. Why should a computer tutor listen? A computer tutor that listens can give help and praise naturally and unobtrusively. We address the following questions: When and how should a computer tutor that listens help students? When and how should it praise students? We examine how the advantages and disadvantages of speech recognition technology helped shape the design and implementation of the Reading Tutor. Despite its limitations, this technology enables the Reading Tutor to provide patient, unobtrusive, and natural assistance for reading aloud.

Keywords

Speech recognition, oral reading, computer-assisted language learning, intelligent tutoring systems, multimedia.

Note: This article is based in part on G. S. Aist and J. Mostow. 1997. Adapting Human Tutorial Interventions for a Reading Tutor that Listens: Using Continuous Speech Recognition in Interactive Educational Multimedia. In *CALL'97 Conference on Multimedia*, Exeter, England.

Giving Help and Praise in a Reading Tutor with Imperfect Listening – Because Automated Speech Recognition Means Never Being Able to Say You're Certain

1. Introduction

Project LISTEN's Reading Tutor uses automatic speech recognition (ASR) to listen to children read aloud and help them. In-school evaluations of successive versions have demonstrated the educational effectiveness of such assistance in helping children read harder material, build fluency, and improve comprehension. These evaluations are reported elsewhere (Mostow et al, 1994; Aist and Mostow, 1997; Mostow and Aist, 1997; Mostow and Aist, in preparation). The present article focuses instead on how the Reading Tutor can make effective use of ASR despite its imperfect accuracy.

In the early grades, word identification skills are a principal bottleneck to fluent reading and comprehension. The design of the Reading Tutor is based in part on interventions used by expert reading teachers to help students learn these skills.

However, human tutors use a range of modalities not available to computers, at least with the robustness and economy required in a school setting. Besides speaking (which computers can also do), human tutors communicate with their students using finger pointing, gaze, gestures, facial expressions, and body language. Conversely, computers surpass human tutors in some respects, such as their ability to dynamically modify a display and to record and replay speech.

How can human tutorial interventions using one set of modalities be adapted to the Reading Tutor, with a different set of modalities? We describe how we have adapted some expert reading interventions for use in the Reading Tutor. Although previous educational software incorporates similar adaptations, the use of continuous speech recognition introduces some novel opportunities and limitations. For example, student attempts at oral reading offer a much richer set of tutorial cues than, say, mouse clicks. These cues enable the Reading Tutor to detect when help is needed, and then engage the student in just-in-time, mixed-initiative spoken dialogue. At the same time, the limited accuracy and speed of speech recognition technology pose some interesting challenges for the design of robust, effective interactions.

2. A Reading Tutor That Listens

Prior Work and Current Implementation

Project LISTEN's automated Reading Tutor builds on the speech analysis methods in Mostow, Roth, Hauptmann, and Kane (1994) and the design recommendations in Mostow, Hauptmann, and Roth (1995). The tutor adapts the Sphinx-II continuous speech recognition system (Huang et al., 1993) as described in Mostow et al. (1994). Unlike its predecessor, the Reading Coach (Mostow et al.), which required a NeXT machine for the user and a Unix workstation for the speech recognizer, the Reading Tutor runs in Windows 95, 98 or NT 4.0 on a Pentium, with a noise-cancelling headset microphone. This platform is cheap enough to put in a school long enough to help children learn to read better. For other research related to using speech recognition to listen to oral reading, see Bernstein and Rtischev (1991), Kantrov (1991), Phillips, McCandless, and Zue (1992), Russell et al. (1996), and IBM (1998).

School Application

The Reading Tutor incorporates materials adapted from *Weekly Reader* (a newsmagazine for children) and other sources. After considerable testing, tuning, and refinement with school children, the Reading Tutor was installed in October 1996 at Fort Pitt Elementary School in Pittsburgh, Pennsylvania. This initial version ran on a single machine in a small room, and was used by a pilot group of third graders under the individual supervision of a school aide. Since 1997, subsequent versions of the Reading Tutor have been used daily in regular classrooms ranging from kindergarten through fifth grade. Students are typically scheduled to read for 10-20 minutes or 1-2 stories per session, depending on the teacher.

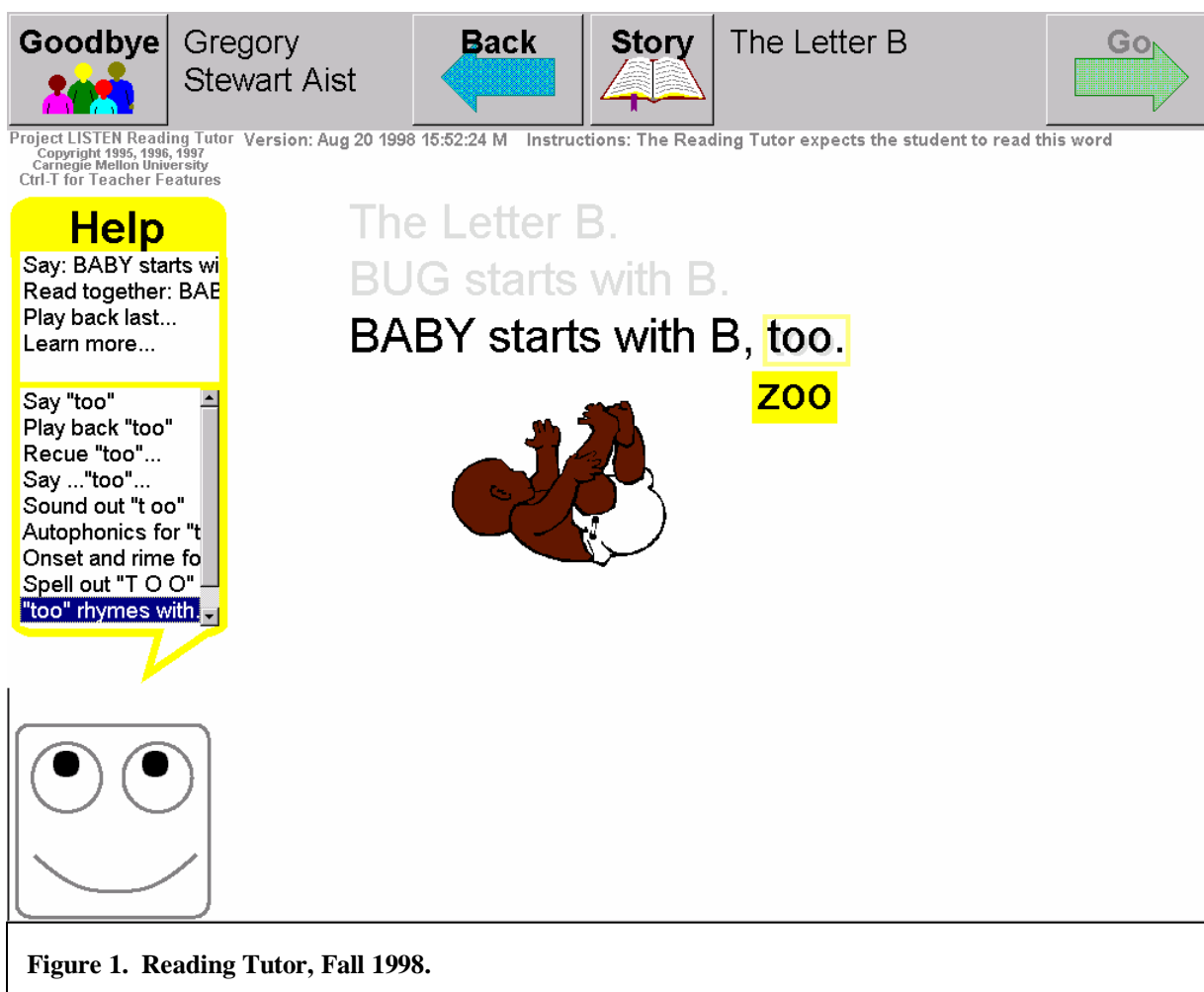


Figure 1. Reading Tutor, Fall 1998.

The Core Interaction: Assisted Reading

The Reading Tutor listens to a child read one sentence at a time, as illustrated in Figure 1. The Reading Tutor displays a simple animated persona that actively watches and patiently listens. The Reading Tutor displays a sentence and may read a difficult word or the entire sentence to the student, highlighting words in yellow as it speaks them.

The Reading Tutor listens to the student read aloud. If it hears the student make a mistake and go on to the next word without self-correcting the mistake, it interrupts by underlining the incorrect word and (sometimes) coughing or “clearing its throat” to catch the student’s attention (Aist, 1998).

When the Reading Tutor hears the end of the sentence or a prolonged silence, it aligns the speech recognizer output against the sentence to decide which words the student read correctly. The Reading Tutor gives the student “credit” for the words it heard the student read correctly, and turns each credited word green to show which words it thinks the student read correctly. When the student has received credit for every important word in the sentence, the Reading Tutor goes on to display the next sentence (Aist, 1997). Otherwise, it responds expressively by using recorded human voices.

Besides speaking the sentence or an individual word, the Reading Tutor may give a rhyming hint or other decoding assistance, or prompt the student to read aloud or click for help. The *Help* balloon lists the menu of assistance available for the sentence and currently selected word. The Reading Tutor selects from this menu when it decides to give help, or when the student clicks on a word. For example, Figure 1 shows the Reading Tutor using the word

“zoo” as a rhyming hint for the word “too.” After giving feedback, the Reading Tutor lets the child reread the word or sentence, or click *Go* to go on to the next sentence.

3. Applying ASR to Oral Reading

Why listen?

The Reading Tutor attempts to address some of the key problems in children’s reading. What are these problems, and how does listening to the student read help the Reading Tutor to address each problem?

One problem is *word identification*: children often misidentify a word or cannot identify it at all. The Reading Tutor addresses word identification by speaking (or giving a hint for) a word that the child gets stuck on, clicks on for help, misreads, or is likely to misread (e.g., has misread in the past, or has never encountered before). Listening to the student is helpful for word identification because young children often lack the metacognitive skills required to realize when they need help, and request it.

A second cognitive problem involves *attentional bottlenecks* (Curtis, 1980). Even when struggling readers eventually identify all the words in a sentence correctly, they often spend so much attention on word identification that they fail to comprehend the sentence. Listening to the student helps address attentional bottlenecks by enabling the Reading Tutor to detect halting, disfluent reading, and respond by reading the sentence aloud, thereby freeing the student to attend to comprehension. Letting the student reread the sentence more fluently also helps comprehension.

A third problem involves *motivation*. One of the best predictors of improvement in reading is time on task. Unfortunately, children who most need practice in reading are often the least inclined to get it, because reading is such a difficult and frustrating process for them. Listening to the student provides powerful motivation by enabling the Reading Tutor to serve as an attentive, perceptive, responsive audience for the student’s efforts. Besides providing spoken and graphical feedback to the student’s reading, listening to the student enables the Reading Tutor to provide an attentive audience in the form of the animated persona displayed in the lower left corner of the screen. The persona gazes at the student’s current position in the sentence, making clear that the Reading Tutor is attending to the student’s reading. As additional visible feedback, the Reading Tutor displays a “shadow” under the word that it thinks the student is trying to read, and turns words green when it accepts them as correct. To reinforce this impression of “active listening,” the Reading Tutor occasionally backchannels (“mm-hmm”, “uh-huh”) when the student hesitates. To project a patient personality, the Reading Tutor listens for appropriately long silences before responding – 2 seconds before backchanneling, 4 seconds before giving a hint, and 7 seconds before prompting the student what to do. Adults often find it difficult to wait so long, but slow readers need patient listeners.

Kinds of Reading Phenomena Detectable by ASR

ASR-enabled listening allows the Reading Tutor to monitor the student’s performance closely. From a functional point of view, the Reading Tutor’s listening capability must meet several real-time computational requirements to serve the tutorial purposes described above.

The Reading Tutor must *detect transitions between silence and speech*. This capability enables the Reading Tutor to detect when the student starts reading, hesitates, or gets stuck.

The Reading Tutor must *track the student’s position in the sentence*. This capability enables the Reading Tutor to visibly follow the student’s reading by subtly shadowing the current word (e.g., the word “too” in Figure 1) and by gazing at it. (When the student moves the mouse, the persona gazes instead at the cursor.) Tracking also enables the Reading Tutor to decide which word to give help on when the student gets stuck. Tracking the student’s position lets the Reading Tutor detect when the student reaches the end of the sentence, so that it can respond promptly instead of waiting for a prolonged silence.

The Reading Tutor must *detect deviations from correct reading*. This capability enables the Reading Tutor to credit correctly read words, provide corrective feedback for mistakes when appropriate, praise good performance, and update its student model of performance on different words. Notice that the current Reading Tutor does not try to phonetically *transcribe* deviations from correct reading. Although such a capability could offer considerable

diagnostic value, accurate phonetic transcription of reading mistakes lies beyond the current state of the art in speech recognition.

How does the Reading Tutor listen?

To achieve the listening capabilities it requires, the Reading Tutor applies ASR to listen to oral reading. Of course not every deviation from perfect reading warrants a tutorial response! The Reading Tutor decides when and how to respond based not only on ASR output but also on pedagogically appropriate criteria.

Using ASR to listen to oral reading involves modeling (however imperfectly) various phenomena of oral reading. Oral reading phenomena include omission, repetition, hesitation, substitution, and insertion. Misreadings include both words and non-words. Word fragments and other speech sounds include false starts, subvocalization, whispering, sounding out, and self-correction. Non-speech sounds include breath noises, microphone bumps, and background noises, among others. In short, there are an infinite variety of deviations from perfect reading.

The ASR engine used by the Reading Tutor is the Sphinx-II speech recognizer (Huang et al, 1993). Given an input utterance, the recognizer searches for the word sequence that matches it the best. Like other modern recognizers, Sphinx-II requires three types of knowledge:

- A set of *acoustic models*, which specify how well a given phoneme (e.g. /S/) matches a given segment of speech signal.
- A *pronunciation lexicon*, which lists a set of known words (e.g. “starts”) and specifies one or more phonemic pronunciations for each word (e.g. /S T AA R TS/).
- A *language model*, which specifies the *a priori* likelihood of different sequences of words (e.g. “starts with”).

In addition, several parameters affect the recognizer in various ways, such as how to weight the quality of acoustic match versus the prior likelihood specified by the language model, and how much memory to use in the search. (More memory can allow greater accuracy.)

To apply ASR to children’s oral reading of a known text, we adapt the approach of Mostow, Hauptmann, Chase, and Roth (1993):

- Use acoustic models trained on adult female speech in other tasks.
- Restrict the lexicon to words in the text.
- Use a language model that at each point (e.g., after the word “Baby” in Figure 1) expects either the correct next word of text (the word “starts”), or (with much lower probability) a jump to some other point in the text.
- Align the ASR output against the text to determine the reader’s position in the text, and to decide which words to accept as correctly read.

Thus the recognizer uses the text words both to represent themselves and as “distractors” to approximate other oral reading phenomena, including out-of-vocabulary words and non-word sounds.

A number of additional ideas help improve speech recognition accuracy for oral reading (Mostow, Roth, Hauptmann, and Kane, 1994):

- Adapt the adult acoustic models using a corpus of children’s read speech.
- Augment the set of distractors for a given text word (e.g., “starts”) to include truncated pronunciations (e.g., /S T AA/), so as to better model false starts, sounding out, and near-misses.
- Restrict the lexicon and language model to just the current sentence, rather than the whole story.
- Ignore mistakes on common function words (e.g. “the”), thereby reducing false alarms (words read correctly but rejected) with little effect on detection of serious reading mistakes.

How well does the Reading Tutor listen?

Mostow, Roth, Hauptmann, and Kane (1994) conducted experiments with the recognizer to adjust the tradeoff between detecting reading mistakes and rejecting correctly read words. This study evaluated the recognizer’s ability to classify each word of text as (eventually) read correctly or not. The test data consisted of children’s oral reading recorded in a Wizard of Oz simulation of an automated “reading coach.” The recognizer detected 49% of the mistakes flagged by the human wizard as serious enough to impair comprehension, with a false alarm rate under 4%. The bias toward lower false alarms was to avoid frustrating students with unnecessary interventions. These results

were achieved using acoustic models trained on adult female speakers. Dr. Alex Hauptmann subsequently achieved further improvements by adaptive training of these acoustic models on a small (12-speaker) corpus of children's oral reading recorded by Dr. Maxine Eskenazi.

However, recognizer accuracy tends to be lower under normal school conditions than in supervised experiments. For example, to function in a noisy classroom environment, the Reading Tutor uses a noise-canceling close-taking headset microphone. This microphone works best when positioned about 2 cm from the speaker's mouth so as to pick up speech and cancel background noise, and slightly below the breath stream so as to avoid breath noise. However, in classroom use it is not uncommon to see children wearing the headset with the microphone positioned too close (touching the lips) or too far (out to the side, several inches away from the mouth). Thus the Reading Tutor's robustness with respect to poor microphone placement may matter more than its accuracy under ideal conditions.

One indication of the Reading Tutor's recognition accuracy under field conditions comes from a study performed for a different purpose (Aist et al., 1998). This study analyzed data recorded by the Reading Tutor in the course of normal in-school use. One finding of this study was that in the subset of sentences that the Reading Tutor classified as perfect (no insertions, substitutions, or deletions), 94% of the words were in fact read correctly. Although this result gives an interesting indication of the Reading Tutor's listening accuracy, it is not comparable with the earlier evaluation, both because of the deliberate selection bias, and because it uses a different measure of accuracy.

Although ASR accuracy can provide a useful measure of the Reading Tutor's listening capability, what really matters is the Reading Tutor's educational effectiveness. For example, when the student misreads multiple words in a sentence, the Reading Tutor may be wrong about which specific words the student missed, yet still respond appropriately by reading the sentence aloud. How can the Reading Tutor behave robustly despite imperfect ASR?

Minimizing the Effects of ASR Errors

Automated speech recognition is less accurate than human hearing. Therefore, the Reading Tutor must behave gracefully even when its speech recognition fails to detect an error or, alternatively, hallucinates an error – a false alarm – when no error exists. What if the tutor is wrong about the student's position in the sentence? We have observed that a human tutor might point at a word, especially if he or she was uncertain about where the student was. The Reading Tutor adapts this strategy to the screen display by using yellow highlighting to show which word it is speaking. If the Reading Tutor is especially uncertain about where the student is, it resorts to position-independent interventions such as reading the sentence.

What if the Reading Tutor is wrong about what words the student read correctly? The tutor never says that the student was right or wrong. Instead, the Reading Tutor responds to a hypothesized incorrect word by modeling the correct word, or it indicates its judgement that the student is incorrect by saying "mmm?"

4. When to Help?

We describe some general types of human tutorial interventions that we have observed or that are part of the emerging literature on tutoring (e.g., Fox, 1993), and we discuss how we have implemented them in the Reading Tutor. We pay particular attention to how the capabilities and limitations of speech recognition, and of computers in general, influence this adaptation. Thus, our design for intervention is shaped by an interaction of pedagogical needs with technical feasibility.

Preempting Students' Mistakes

Elementary school teachers often preview difficult vocabulary with students before assigning them a story. The intent of preemptive assistance is to prevent mistakes before they occur. Preventing a mistake is generally more effective than correcting it (Hebb, 1949). The Reading Tutor provides preemptive assistance by reading a word or supplying other word-based help immediately upon displaying the sentence and before the student starts reading. First, the Reading Tutor estimates the difficulty of each word in the sentence based on (a) the student's performance on that word in the past and (b) the length of the word. Then, the Reading Tutor identifies one of the most difficult

words in the sentence, and gives help on it. The Reading Tutor is able to adapt this sort of help to an individual student because of the ability of computers to record and analyze large amounts of data on student performance. In the specific ability to record and track which words a learner had previous difficulty with, a computer tutor may even surpass human tutors.

Providing Hints

Human tutors provide hints to their students – partial information about the correct answer that is supposed to help the student discover the answer independently. Traditional hints work in part because the human tutor can judge the correctness of the student’s answer. The Reading Tutor’s speech recognition is not perfect. The Reading Tutor’s initial version of rhyming hints simply highlighted the word in question (e.g., “frog”) and spoke the hint (e.g., “This word rhymes with ‘dog.’”) However, we found that students would often repeat “dog” instead of guessing “frog” – potentially mislearning a word. A human tutor would hear and correct such an error immediately. But the Reading Tutor’s hearing is too limited to detect this difference as accurately as we would like. To prevent this misunderstanding, we modified the Reading Tutor to show the text of the rhyming hint beneath the word, so as not to confuse the hint with the text word. Figure 1 shows the word “zoo” given as a rhyming hint for the text word “too”.

Interrupting When an Error Is Not Self-Corrected

Human tutors exhibit a strong preference for allowing self-correction (Fox, 1993). With the Reading Tutor’s predecessor, the reading coach (Mostow et al., 1994), we observed frequent self-correction by students. The Reading Tutor interrupts only when it appears the student has made an error on a word but not corrected it before moving on to the next word. Our goal is that, if the position estimate is correct, the interruption should draw the student’s attention; if the position estimate is incorrect, the interruption should not be disruptive. Speaking the correct word could be disruptive if the Reading Tutor were wrong about which word the student was reading. Therefore, the Reading Tutor highlights the word and (sometimes) “coughs” or “clears its throat” (by playing an appropriate recorded sound file) to subtly call the student’s attention to the missed or incorrect word.

Responding To Requests For Help

Rather than trying to enable the Reading Tutor to respond to oral requests for help, we have reserved speech input for reading. This makes the speech recognition task easier because there are fewer things the student might say, and it sets up the expectation that “every time you speak, the computer thinks you’re reading” which may serve to limit “off-task” or non-reading speech. Therefore, the Reading Tutor responds to requests for help that students make by clicking on words, under the sentence, or on the Help balloon.

Providing Corrective Feedback

Human tutors provide corrective feedback to a student by, for example, supplying words that the student missed. The Reading Tutor decides when to provide corrective feedback much like its predecessor, the reading coach (Mostow et al., 1994). If the student misses a single content word, the Reading Tutor gives corrective feedback on it. If the student misses multiple words, the Reading Tutor may give help on the first one, or read the entire sentence. The reading coach sometimes expected the student to reread individual words, and other times to read the entire sentence, and these alternating expectations were confusing. The Reading Tutor avoids such confusion by always allowing complete reading as an acceptable response. In brief, the Reading Tutor compares the words output by the speech recognizer to the words of the current sentence; if important words were missed, it either provides help on an individual word, or reads the entire sentence.

Prompting the Student

We assume that a long period of silence indicates that the student needs help. If the Reading Tutor does not detect speech for more than seven seconds, it either prompts the student to read the sentence, or speaks the sentence itself and then encourages the student to read it. While this strategy works for students still at the oral reading stage, it does not extend well to silent reading.

5. How to Help? Modeling Subdialogues of Human Tutors

Human tutors can engage students in subdialogues. For example, a human tutor can ask a student who is reading to sound out a troublesome word, can listen to him or her sound the word out, and can provide feedback on his or her attempt. Subdialogues have been incorporated to good effect in intelligent tutoring systems and learning environments (e.g., Wenger, 1987). But the Reading Tutor's hearing is imperfect, and including subdialogues makes the interaction more brittle. To ensure robust interaction, the task is always to read the current sentence. Therefore, the Reading Tutor *models* common interventions for the student instead of *prompting* the student to do them and *monitoring* the results.

What types of interventions can the Reading Tutor model for the student?

- The Reading Tutor can *read the sentence* to the student. Reading the sentence is expensive for human teachers, but the Reading Tutor can do it cheaply, repeatedly, and patiently. The Reading Tutor plays a recording of the sentence and highlights each word as it is spoken. If no recording is available, the Reading Tutor reads the sentence one word at a time using the individual word recordings. If an individual word recording is not available, the Reading Tutor uses synthesized speech for that word.
- The Reading Tutor can *recue* a word by playing the words leading up to that word, and underlining the word. The intent of *recue* is to put the student back in the context of the sentence before attempting the word again.
- The Reading Tutor can *play back* the student's last recording for a word or sentence. The Reading Tutor cannot be sure that the student read correctly, so it provides this response only when the student clicks on "play back" in the *Help* balloon. Consequently such playback is seldom used except for our own debugging.
- The Reading Tutor can *supply a word* by playing a recording of that word. For homographs, the Reading Tutor's recordings include both pronunciations: "PREsent or preSENT."
- The Reading Tutor can *supply a context-specific reading of a word* by playing the portion of the sentence recording that corresponds to that word. The Reading Tutor thus provides an easy solution to disambiguating homographs. The narration contains the correct in-context pronunciation of the homograph: "Mary bought Bob a PREsent."
- The Reading Tutor can *sound out* a word by pronouncing each phoneme while displaying the corresponding letter(s). The Reading Tutor can also *syllabify* a word by pronouncing the word's syllables while displaying the corresponding letter(s). To pronounce a syllable for which it lacks a recording, the Reading Tutor pronounces the individual phonemes that make up the syllable. The Reading Tutor can similarly supply the *onset and rime* for a word, for example saying and highlighting "d" (/D/) and then "og" (/AH G/) for "dog".
- The Reading Tutor can provide "*autophonics*" assistance by picking out the letter-to-sound correspondence that it thinks the student is having the most trouble with, and supplying just that correspondence. For example, if the word is "dog", the Reading Tutor might highlight the "g", and say /G/.
- The Reading Tutor can *spell* a word by saying each letter while displaying the letters in sequence. This intervention aims at young readers who haven't finished mastering the alphabet, careless readers who need to look more closely at how the word is spelled, and dyslexic readers who may reverse letters, e.g., "d" as "b".
- The Reading Tutor can give a *rhyming hint* for a word by supplying a hint whose end sounds and is spelled the same as the word. For example, "curious" might be supplied as a rhyming hint for "furious" because their ends sound the same and are spelled the same as well. The Reading Tutor's use of precomputed tables enables it to match or exceed human performance at generating rhyming hints in real-time.
- The Reading Tutor can supply a hint that *starts the same* as the word, such as "dog" as a hint for "dogs".

6. When To Praise?

Using automatic speech recognition means never being able to say you're certain. Therefore both corrective and confirmatory feedback must be phrased to avoid explicitly stating that the student was wrong (or right). Mostow et al. (1994) describe how to finesse this issue at the level of individual words – instead of saying whether the student was right or wrong, just echo the correct word, and let the student decide whether to interpret this feedback as confirmatory or corrective. However positive reinforcement is still important motivationally, so it's not enough to read correctly to the student. Our solution is to reinforce larger units of performance than individual words.

Praising Achievement

If the Reading Tutor analyzes the student's reading as a word-by-word perfect reading of the sentence, it sometimes provides positive feedback such as "Excellent." Random positive reinforcement is well known to be more effective than constant reinforcement; also, providing positive feedback for every correctly read sentence would quickly annoy a good reader. The Reading Tutor is not always correct about the student's performance. However, even if the student misread some words he or she may have read much of the sentence correctly. Providing positive reinforcement at the sentence level instead of at the word level thus compensates for inaccurate hearing by the Reading Tutor.

At the end of a story, the Reading Tutor always provides praise. Even if the student did not in fact read very well, the unit of performance is large enough that students will get this praise only 2-3 times per session. Furthermore, praising effort (and not just performance) is acceptable.

Praising Improvement

The Reading Tutor measures fluency by looking at the student's accuracy (percentage of words read correctly) and the inter-word latency between successfully read text words (Mostow and Aist, 1997). When the student's last attempt was more fluent than the previous attempt, the Reading Tutor sometimes says something encouraging, such as: "You're catching on."

7. How To Praise?

The Reading Tutor praises student performance to reinforce success. The Reading Tutor also praises the student to support motivation and self-confidence. Motivational effectiveness is fiendishly difficult to assess, but we have noticed that even skilled adult readers appear to enjoy praise from the Reading Tutor.

Praising the Performance

Some of the phrases that the Reading Tutor uses to praise the student are "Good," "Excellent," and other phrases directed at the student's performance. Praising performance is intended to reinforce success.

Praising the Student

Other phrases are directed at the student. For example, the Reading Tutor may say "You're a good reader" or "You're catching on." Praising the student is intended to support student motivation and self-confidence.

8. Summary

Project LISTEN's Reading Tutor listens to children read aloud and helps them through difficulties.

- Why listen? Listening is important so that computer tutors can give help and give praise based on observation of spoken language use. Automatic speech recognition (ASR) is the basic technology to enable a listening tutor. However, equally critical is adapting the human interface and pedagogy to work with ASR.

- When should a computer tutor help? Helping before the student reads lets the Reading Tutor prevent mistakes before they occur. Helping while the student is reading lets the Reading Tutor encourage the student to continue when stuck and to self-correct mistakes. Helping after the student reads by giving corrective feedback lets the Reading Tutor correct any mistakes that the student did not self-correct.
- How should a computer tutor that listens help students? Because conducting dialogues with subtasks is difficult, the task with the Reading Tutor is always to read all or part of the sentence. The Reading Tutor presents a sentence, assists the student in reading the sentence, and then demonstrates ways of reading words successfully.
- When should a computer tutor praise? The Reading Tutor praises both achievement, when performance is good, and improvement, when performance gets better.
- How should a computer tutor praise? The Reading Tutor praises both the performance and the student.

The Reading Tutor is built around using continuous speech recognition to follow children's oral reading. Its design compensates for inaccurate speech recognition. By using speech recognition, the Reading Tutor is able to give help and give praise naturally and unobtrusively. Thus despite the limitations of ASR, listening helps lower the barrier between the student and the Reading Tutor.

Acknowledgements

This material is based upon work supported in part by the National Science Foundation under Grant No. IRI-9505156 and CDA-9616546, by the Defense Advanced Research Projects Agency under Grant Nos. F33615-93-1-1330 and N00014-93-1-2005, and by the second author's National Science Foundation Graduate Fellowship and Harvey Fellowship. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the official policies, either expressed or implied, of the sponsors or of the United States Government. We thank the Principal of Fort Pitt Elementary School, Dr. Gayle Griffin, and the teachers at Fort Pitt for their assistance; Drs. Rollanda O'Connor and Leslie Thyberg for their expertise on reading; Raj Reddy and the CMU Speech Group (especially Ravi Mosur) for the Sphinx-II speech recognizer; Dan Barritt, Kerry Ishisaki, Rebecca Kennedy, DeWitt Latimer IV, Bryan Nagy, and David Sell, for helping implement the improvements described here; and the many past members of Project LISTEN whose work contributed to previous versions of the Reading Tutor, especially Alex Hauptmann, Steve Roth, Lin Chase, Bob Weide, Matthew and Lee Ann Kane, and Adam Swift for their work on the reading coach that preceded the Reading Tutor; and many students, educators, and parents for tests of the Reading Tutor in our lab and at Fort Pitt Elementary School.

References (see also www.cs.cmu.edu/~listen)

- Aist, G. S. (1998). Expanding a time-sensitive conversational architecture for turn-taking to handle content-driven interruption. *Proceedings of the Fifth International Conference on Speech and Language Processing (ICSLP98)*, Sydney, Australia.
- Aist, G. S. (1997). Challenges for a mixed initiative spoken dialog system for oral reading tutoring. Paper presented at the Symposium on Computational Models for Mixed Initiative Interaction, American Association for Artificial Intelligence Annual Meeting, Palo Alto, CA.
- Aist, G., Chan, P., Huang, X. D., Jiang, L., Kennedy, R., Latimer, D., Mostow, J., and Yeung, C. (1998). How effective is unsupervised data collection for children's speech recognition? *Proceedings of the Fifth International Conference on Speech and Language Processing (ICSLP98)*, Sydney, Australia.
- Aist, G. S., and Mostow, J. (1997). When speech input is not an afterthought: A reading tutor that listens. *Proceedings of the Workshop on Perceptual User Interfaces*, Banff, Canada.
- Bernstein, J., & Rtischev, D. (1991). A voice interactive language instruction system. *Proceedings of the Second European Conference on Speech Communication and Technology (EUROSPEECH91)*, Genova, Italy.

- Curtis, M.E. (1980). Development of components of reading skill. *Journal of Educational Psychology*, 72(5), 656-669.
- Fox, B. A. (1993). *The human tutorial dialogue project: Issues in the design of instructional systems*. Hillsdale, NJ: Lawrence Erlbaum.
- Hebb, D. O. (1949). *The organization of behavior*. New York: John Wiley and Sons.
- Huang, X. D., Alleva, F., Hon, H. W., Hwang, M. Y., Lee, K. F., and Rosenfeld, R. (1993). The Sphinx-II speech recognition system: An overview. *Computer Speech and Language* 7(2):137-148.
- IBM. 1998. Watch Me Read. <http://www.ibm.com/IBM/IBMGives/k12ed/watch.htm>.
- Kantrov, I. (1991). Talking to the computer: a prototype speech recognition system for early reading instruction. Technical Report 91-3. Center for Learning, Teaching, and Technology, Education Development Newton, MA.
- Mostow, J., Hauptmann, A. G., Chase, L. L., and Roth, S. F. (1993). Towards a reading coach that listens: Automated detection of oral reading errors. *Proceedings of the Annual Meeting of the American Association for Artificial Intelligence (AAAI 93)*, Washington, DC.
- Mostow, J., Roth, S. F., Hauptmann, A. G., & Kane, M. (1994). A prototype reading coach that listens. *Proceedings of the Annual Meeting of the American Association for Artificial Intelligence (AAAI 94)*, Seattle, WA.
- Mostow, J., Hauptmann, A., & Roth, S. F. (1995). Demonstration of a reading coach that listens. *Proceedings of the Eighth Annual Symposium on User Interface Software and Technology*, Pittsburgh PA. (Sponsored by ACM SIGGRAPH and SIGCHI in cooperation with SIGSOFT).
- Mostow, J., and Aist, G. S. (1997). The sounds of silence: Towards automatic evaluation of student learning in a reading tutor that listens. *Proceedings of the Annual Meeting of the American Association for Artificial Intelligence (AAAI 97)*, Palo Alto, CA.
- Mostow, J., and Aist, G. S. (in preparation). Evaluating tutors that listen: An overview of Project LISTEN. To appear in K. Forbus and P. Feltovich (Eds.), as-yet untitled collection on artificial intelligence and education. Palo Alto, CA: AAAI Press.
- Phillips, M., McCandless, M., & Zue, V. (1992). Literacy tutor: An interactive reading aid. Technical Report, Spoken Language Systems Group, Laboratory for Computer Science, Massachusetts Institute of Technology.
- Russell, M., Brown, C., Skilling, A., Series, R., Wallace, J., Bonham, B., & Barker, P. (1996). Applications of automatic speech recognition to speech and language development in young children. *Proceedings of the Fourth International Conference on Spoken Language Processing*, Philadelphia PA.
- Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems*. Los Altos, CA: Morgan Kaufman.

Biodata:

Jack Mostow is founder and director of Carnegie Mellon University (CMU)'s Project LISTEN, which is getting computers to listen to children read aloud. He is also Senior Research Scientist with the CMU Robotics Institute, Language Technologies Institute, and Human-Computer Interaction Institute. His Ph.D. is in computer science from CMU, and he has held research and faculty positions at Stanford University, the University of Southern California's Information Sciences Institute, and Rutgers University. His Project LISTEN work earned the Outstanding Paper Award at the Twelfth National Conference on Artificial Intelligence in August 1994. Dr. Mostow has authored several dozen publications in the field of artificial intelligence, and has served as an Editor of Machine Learning Journal and of IEEE Transactions on Software Engineering.

Jack Mostow
215 Cyert, CMU-LTI,
4910 Forbes Ave.
Pittsburgh, PA 15213-3702

412-268-1330 voice; -6298 FAX

www.cs.cmu.edu/~listen

Greg Aist is a Ph.D. student in the Language Technologies Institute (LTI) at Carnegie Mellon University. He holds a National Science Foundation Graduate Research Fellowship and a Harvey Fellowship. His research interests are in human-computer spoken dialog, computational linguistics, and computer-assisted language learning.

Greg Aist
LTI, CMU, 4910 Forbes Avenue
Pittsburgh, PA 15213-3734
412 268 8496 voice; -6298 FAX
aist@andrew.cmu.edu