Adapting Human Tutorial Interventions for a Reading Tutor that Listens: Using Continuous Speech Recognition in Interactive Educational Multimedia

Gregory Aist and Jack Mostow
Project LISTEN
Carnegie Mellon University
4910 Forbes Avenue
Pittsburgh, PA 15213-3734
+1 412 268 8496
aist@andrew.cmu.edu

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 listens to children read aloud, and helps them. Why should a computer tutor listen? A computer tutor that listens can give help and give praise naturally and unobtrusively. In this paper, we address the following questions: When and how should a computer tutor that listens help students? When and how should a computer tutor that listens praise students? We examine how the advantages and disadvantages of speech recognition helped shape the design and implementation of the Reading Tutor. Despite its limitations, speech recognition enables the Reading Tutor to provide patient, unobtrusive, and natural assistance for reading out loud.

Keywords

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

INTRODUCTION

Project LISTEN's Reading Tutor listens to children read aloud, and helps them. In-school evaluations of successive versions have demonstrated the Tutor's ability to assist students' fluency and comprehension. Word identification skills are a principal bottleneck to fluent reading and comprehension in the early grades. 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 some 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. However, the limited accuracy and speed of speech recognition technology pose some interesting challenges for the design of robust, effective interactions.

A READING TUTOR THAT LISTENS

Our automated Reading Tutor builds on the speech analysis methods in (Mostow et al. 1994) and the design recommendations in (Mostow et al. 1995). Unlike the reading coach in (Mostow et al. 1994), which required a NeXT machine for the user and a Unix workstation for the speech recognizer, the Reading Tutor runs in Windows™ 95 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. The Tutor incorporates materials adapted from *Weekly Reader* (a newsmagazine for children) and other sources. 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).

The Reading Tutor was installed in August 1997 at Fort Pitt Elementary School in Pittsburgh, Pennsylvania, U.S.A. Eight classrooms (one kindergarten, one first grade, one second grade, three third grade, and two fourth grade) have one Reading Tutor each. Children read with the Reading Tutor for approximately twenty minutes per session.

Figure 1. Reading Tutor, Fall 1997.

The Core Interaction: Assisted Reading

The Reading Tutor (Figure 1) listens to a child read one sentence at a time. 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. Then, the Reading Tutor listens to the student read aloud. If the Reading Tutor hears the student make a mistake and does not hear the student self-correct that mistake, it interrupts by flashing the incorrect word and (sometimes) coughing or "clearing its throat" to catch the student's attention. 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. 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, the Reading Tutor responds expressively by using recorded human voices (Aist and Mostow 1997). For example, the Reading Tutor may speak a word or an entire sentence. After giving feedback, the Reading Tutor lets the child reread the word or sentence. Besides reading words or sentences, the student can use the mouse to get help on a word or sentence.

WHY LISTEN?

Listening allows the Reading Tutor to closely monitor the student's performance. The Reading Tutor's listening capability also provides powerful motivation for the student. The Reading Tutor adapts the Sphinx-II continuous speech recognition system (Huang et al. 1993) as described in (Mostow et al. 1994). The Reading Tutor uses speech recognition to track the student's position in the sentence. The Reading Tutor also uses speech recognition to decide which words the student read correctly.

Automated speech recognition is less accurate than human hearing. Therefore, the Reading Tutor must behave gracefully even when its speech recognition is incorrect.

What if the Reading Tutor is wrong about the student's position in the sentence? A human tutor might point at a word, especially if he or she was uncertain about where the student was. The Reading Tutor uses yellow highlighting to establish focus. 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 Reading Tutor never says that the student was right. It also never says that the student was wrong. Instead, the Reading Tutor responds by modeling the correct word, or indicates its judgement by saying "mmm?" when it believes the student is incorrect.

WHEN TO HELP?

We describe some general types of human tutorial interventions and how we have implemented them in the Reading Tutor. We pay particular attention to how the capabilities and limitations of speech recognition, and computers in general, influence this adaptation.

Preempting Mistakes

Elementary school teachers often preview difficult vocabulary before assigning a story to students. The intent of preemptive assistance is to prevent mistakes before they occur. Preventing a mistake is much 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. First, the Reading Tutor estimates the difficulty of each word in the sentence based on the student's performance on that word in the past, and on the length of the word. Then, the Reading Tutor identifies the two most difficult words in the sentence. If both are "hard" words, the Reading Tutor randomly chooses one of those words 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.

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. Originally, the Reading Tutor provided hints such as "This word rhymes with *dog*" (for a word such as *frog*). We found that students would repeat *dog* instead of guessing *frog* – potentially mislearning a word. The Reading Tutor's hearing is too limited to detect this difference as accurately as we would like. We have therefore modified the traditional concept of "hint" to be "hint, pause, then supply answer" so that for now the Tutor always provides the correct word.

Interrupting When An Error Is Not Self-Corrected

Human tutors exhibit a strong preference for allowing self-correction (Fox 1994). With the Reading Tutor's predecessor, the reading coach, 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. Speaking the correct word could be disruptive if the Reading Tutor was wrong about which word the student was reading. 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. Therefore, the Reading Tutor highlights the word and (sometimes) "coughs" or "clears its throat" (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 verbal requests for help, we have reserved the speech input channel 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 made by clicking on words 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). However, unlike the reading coach, the Reading Tutor only gives corrective feedback on single words or the entire sentence, not two words at a time. In brief, the Reading Tutor compares the words output by the speech recognizer to the words of the current sentence, and if important words were missed either provides help on an individual word or reads the entire sentence.

Prompting The Student

We assume a long period of silence indicates 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.

HOW TO HELP?

Human tutors can engage students in subdialogues. For example, a human tutor can ask a student to sound out a word, listen to him or her sound the word out, and provide feedback on his or her attempt. 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 help 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. To read the sentence, 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 will read 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 *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 corresponding to the 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, the Reading Tutor currently pronounces the individual phonemes that make up the syllable.

The Reading Tutor can *spell* a word by saying each letter while displaying the letters in sequence.

Additional phonologically based feedback, such as supplying a *rhyming word* or a *word that starts the same* have been implemented and tested with children. These features are not currently included in the installation of the Reading Tutor at Fort Pitt Elementary because of excessive memory requirements.

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 stating that the student was wrong (or right). We describe elsewhere how we finesse this issue at the level of individual words (Mostow et al. 1994). But positive reinforcement is still important motivationally, so we 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." 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 the word level thus compensates for the Reading Tutor's inaccurate hearing.

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 only get this praise 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 inter-word latency (the gaps of silence between 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." Providing positive feedback for every correctly read sentence would quickly annoy a good reader.

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.

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.

CONCLUSION

Project LISTEN's Reading Tutor listens to children read aloud, and helps them. Why listen? Listening is important so that computer tutors can give help and give praise based on observation of spoken language use.

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 encourage the student 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. The Reading Tutor's design compensates for inaccurate speech recognition. By using speech recognition, the Reading Tutor is able to give help and give praise naturally and unobtrusively. Speech recognition helps lower the barrier between the student and the Reading Tutor.

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

1. Aist, G. S. 1997. Challenges for a Mixed Initiative Spoken Dialog System for Oral Reading Tutoring. AAAI 1997 Spring Symposium on Computational Models for Mixed Initiative Interaction.

2. Bernstein, J., and Rtischev, D.. 1991. A voice interactive language instruction system. Proceedings of the Second European Conference on Speech Communication and Technology (EUROSPEECH91). Genova, Italy.

3. Fox, B. A. 1993. The Human Tutorial Dialogue Project: Issues in the Design of Instructional Systems. Hillsdale NJ: Lawrence Erlbaum.

4. Hebb, D. O. 1949. The Organization of Behavior. New York: John Wiley and Sons.

5. 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.

6. 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 Center, 55 Chapel Street, Newton, MA 02160.

7. Mostow, J., Roth, S. F., Hauptmann, A. G., and Kane, M. 1994. A Prototype Reading Coach that Listens. In Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94), Seattle WA.

8. Mostow, J., Hauptmann, A., and Roth, S. F. 1995. Demonstration of a Reading Coach that Listens. In Proceedings of the Eighth Annual Symposium on User Interface Software and Technology, Pittsburgh PA. Sponsored by ACM SIGGRAPH and SIGCHI in cooperation with SIGSOFT.

9. Mostow, J., and Aist, G. S. 1997. The Sounds of Silence: Towards Automatic Evaluation of Student Learning in a Reading Tutor that Listens. To appear in the 1997 National Conference on Artificial Intelligence (AAAI 97).

10. Phillips, M., McCandless, M., and V. Zue. September 1992. Literacy Tutor: An Interactive Reading Aid Technical Report, Spoken Language Systems Group, MIT Laboratory for Computer Science, MIT.

11. Russell, M., Brown, C., Skilling, A., Series, R., Wallace, J., Bonham, B., and 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.