When Speech Input is Not an Afterthought: A Reading Tutor that Listens

Jack Mostow and Gregory Aist
Project LISTEN, Carnegie Mellon University
CMU-LTI, 215 Cyert Hall, 4910 Forbes Avenue, Pittsburgh, PA 15213-3734
mostow@cs.cmu.edu

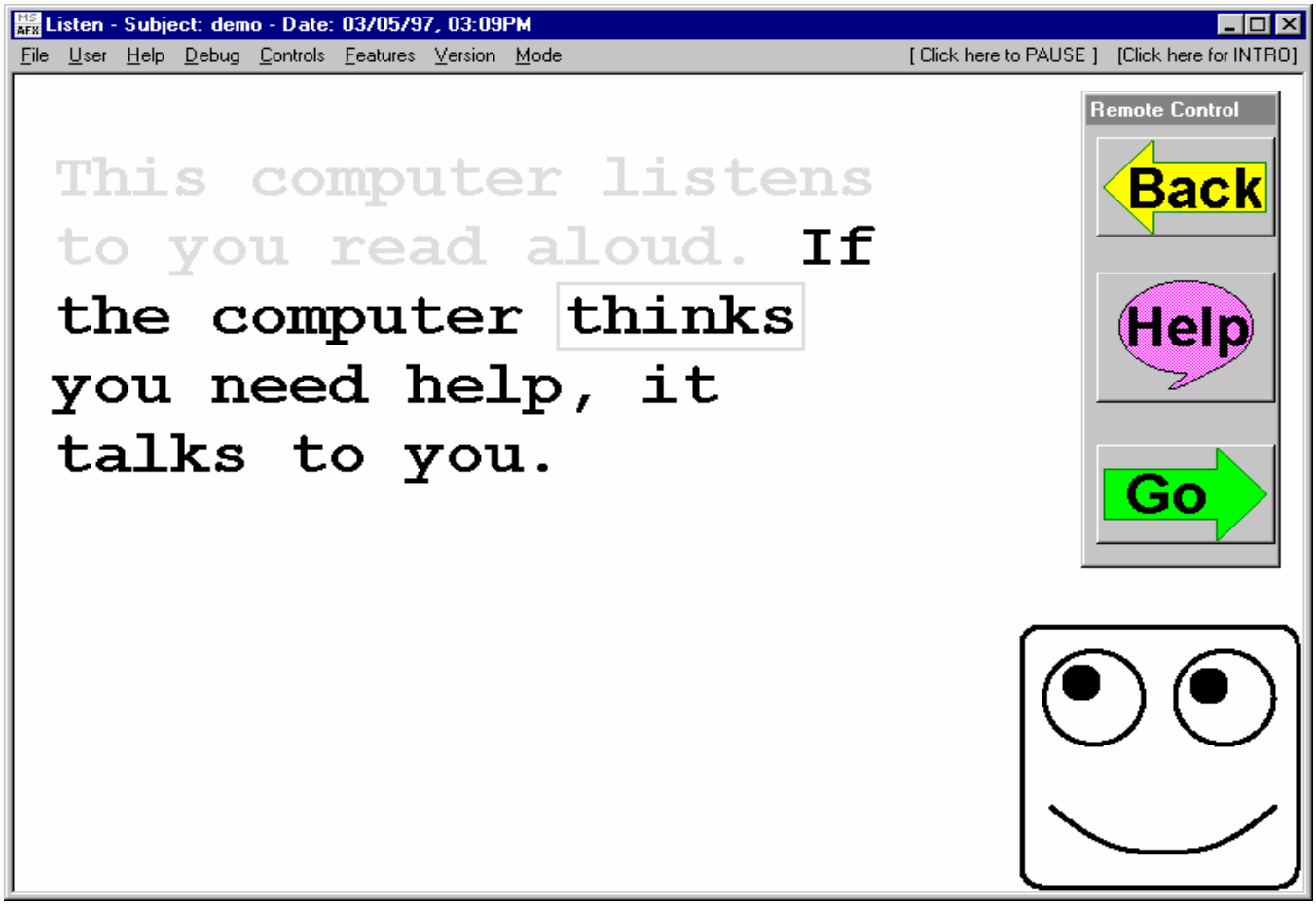
Abstract

Project LISTEN's Reading Tutor listens to children read aloud, and helps them. The first extended in-school use of the Reading Tutor suggests that for this task speech input can be natural, compelling, and effective.

1. Introduction

The "retrofit" approach to integrating perceptual input into the user interface treats the new input as an add-on to improve an existing interface. This "add-on" approach is more or less limited to tasks amenable to the existing interface, which it implicitly assumes will remain at the core of the design. This approach can therefore fail to produce a substantially more natural or compelling interface. For example, voice-driven menus have failed to gain user acceptance by office workers who are reluctant to speak to their computers.

In contrast, Project LISTEN's Reading Tutor is based on a user interface designed from scratch around speech input for a task where it is essential – individual tutoring for oral reading. This "outside-in" design process, and the Wizard of Oz experiments it used, were reported in [Mostow et al, AAAI94]. Studies of the Reading Tutor's predecessors had evaluated their usability and assistive effectiveness, but were based on at most an hour or two of use per subject, conducted in the presence of research personnel. Here we report on the first experiment to examine the effects of extended in-school use without the researchers present. A screenshot of the November 1996 version of the Reading Tutor used in this experiment is shown below.



2. Background

Project LISTEN is an inter-disciplinary research project at Carnegie Mellon University to develop a novel tool for literacy -- an automated Reading Tutor that displays stories on a computer screen, listens to the student read aloud, and responds with spoken and graphical assistance.

The Reading Tutor adapts Carnegie Mellon's state-of-the art Sphinx-II speech recognizer to analyze the student's oral reading [Mostow et al, AAAI93, AAAI94] so as to track the current position in the text and identify reading mistakes. The Reading Tutor's responses to oral reading are modelled in part after expert reading teachers, but adapted to the strengths and limitations of automated speech recognition [Aist and Mostow, CALL97].

In October 1996 we deployed the implemented portion of the Reading Tutor on a 64 MB, 90 MHz Pentium in an urban Pittsburgh elementary school for the first test of extended daily use. The purpose of this pilot experiment was to explore how the Tutor could help children learn over time to read better. Would the Tutor be robust enough to operate for extended periods in a school environment without crashing? Would children be willing to continue using the Reading Tutor after the novelty wore off? What other issues would arise? Most important, would using the Reading Tutor improve their reading?

The pilot group, chosen by the school, consisted of the eight lowest-reading third graders (the two poorest readers from each of the four third-grade classrooms). These children were considered at greatest risk of growing up illiterate. According to individual assessments administered in November 1996 by the school district's reading diagnostician, the children started out 2-3 years below grade level. Most of these students had stayed with the same teachers since kindergarten and had made almost no progress in reading since then.

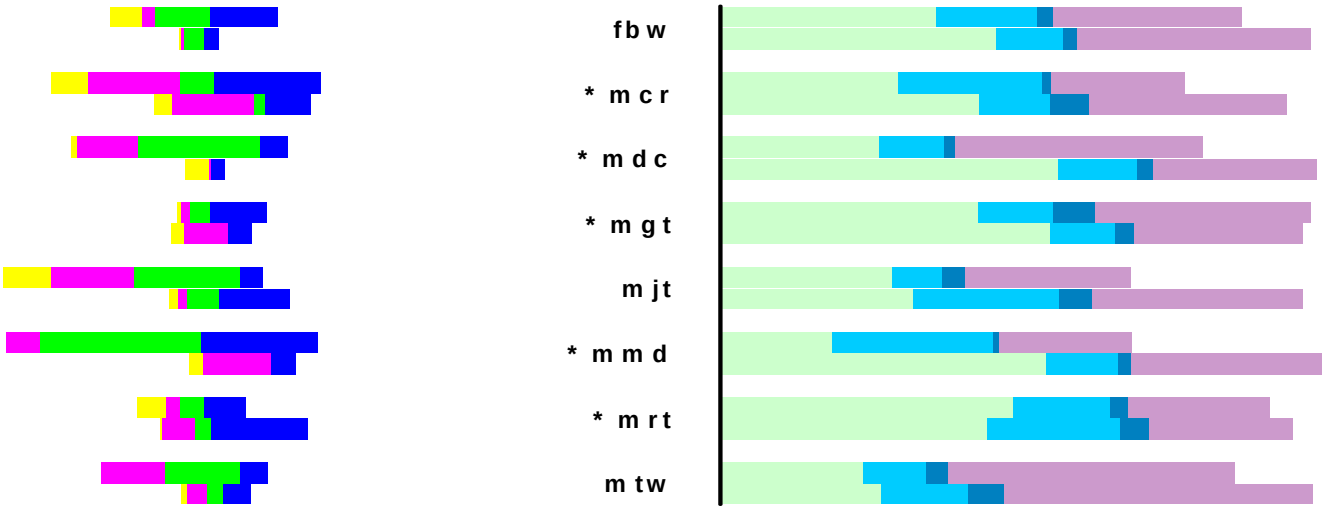
A school aide escorted each subject to and from class to use the Reading Tutor in a small room. The aide helped the student choose from a menu of text-only stories adapted from *Weekly Reader* and other sources. Children had favorite stories, which they often reread.

Based on initial pilot use, we made some additional modifications to the Reading Tutor, mostly to reduce dialogue breakdowns caused by speech recognizer errors [Aist & Mostow, CMMII 97]. We froze the code with the version of November 7, 1996, and kept the same version for the rest of the pilot study, other than adding some stories about halfway through when the most voracious readers started running out of text. The pilot experiment concluded on June 12, 1997 at the end of the school year.

3. Pilot Version of Reading Tutor

The Reading Tutor takes speech and mouse input, and emits speech, text, and graphical output. The Reading Tutor displays the story text incrementally, adding one sentence at a time for the child to read aloud. The user may read aloud, click on a word to have the Reading Tutor speak it, click on the Help balloon to hear the sentence, or click on a Go or Back button to move to the next or previous sentence [Mostow et al, UIST95; Mostow and Aist, AAAI97]. A simple graphical persona simulates an animate listener by gazing at the current word or blinking.

The version of the Reading Tutor used for this pilot experiment employed a limited set of responses. The November 1996 Reading Tutor responded when it detected the end of the sentence, a 4-second hesitation, or a button click. If the Reading Tutor accepted the reading, or if the child clicked the "Go" button, it went on to the next sentence. If the Reading Tutor detected a single missed word, it highlighted the word in pink and spoke it, or



recued it by rereading the words that led up to it. If the Reading Tutor detected more than one missed word, it highlighted the missed words and read the entire sentence aloud. Then the child could try the sentence again.

4. Ease of Use

How easy is the Reading Tutor to use? We found from the pilot study and other experience that children of varying ages (5 and up) and computer experience (including utter novices) begin using the Reading Tutor with less than two minutes of training. The Reading Tutor is simple enough to use that children have been able to teach other children how to use it.

Most initial training covers how to put on the headset microphone and use the mouse. Children can then use the Reading Tutor to help them read a short tutorial covering other aspects. This "bootstrapping" scheme for user training is made feasible by "just-in-time" spoken assistance provided by the Reading Tutor in response to prolonged silence on the part of the user. This assistance takes the same form for any text read by the student, not just the tutorial. The assistance consists mainly of reading aloud to the student, but also prompts the student on what to do, such as read the sentence or click the Go button to go on.

The Reading Tutor responds to mouse clicks, but is designed to allow hands-free operation for reading a story. Thus the user can choose whether to read (and rely on the Reading Tutor to help or go on), or use the mouse. To characterize different patterns of usage and divisions of labor between student and Reading Tutor, we computed the distributions of student and Reading Tutor actions in the bar chart above. (This chart is rather complex, and monochrome hardcopy may be less clear than viewing it in color on-line). The chart is based on the 4202 actions logged by the Reading Tutor in November 96 and the 4033 actions in February 97. The top bar for each subject is for November 1996, and the bottom bar is for February 1997. The starred (*) subjects gained significantly in fluency over that period.

The successive segments in the left side of each bar indicate the relative frequency of different student actions: click Back, Help, Go, or a word, or read aloud. Conversely, the right side shows the distribution of Reading Tutor actions: go on to the next sentence, recue a word, read a word, or read the sentence. Thus the horizontal position of each bar reflects the relative number of student versus Tutor actions. For the ideal case of perfect reading and speech recognition, the only student actions would be to read sentences, and Tutor actions would consist solely of advancing to the next sentence.

The chart reveals large differences in usage patterns between subjects and over time. In particular, 6 of the 8 students in the pilot study came to rely more on speech, clicking much less often in February than in November.

5. Motivation

How compelling is the Reading Tutor? Interviews with the students and their teachers showed continued strong interest and motivation on the part of the children in using the Reading Tutor right up through the last week of the school year, even after nearly eight months of use. The teachers attributed the children's changes in classroom attitude and reading performance to the Reading Tutor, and reported that the children looked forward eagerly to their daily sessions with it.

What was responsible for this dramatic motivational effect? Motivation is complex, context-sensitive, and problematic to measure. However, it is instructive to point out some factors that could *not* be responsible for children's reported motivation to use the Reading Tutor. The Reading Tutor lacked the flashy graphics and sound effects of commercial software and video games. It did not even include pictures – just text, which [Malone 1981] found was the least interesting element of computer games. Moreover, reading was especially difficult and frustrating for the pilot subjects. We believe that the Reading Tutor's novel ability to listen is essential to its motivational power.

6. Effectiveness

Student progress was evaluated in two ways. As pre- and post-tests to measure changes in individual reading ability, school reading specialists administered individual reading assessments. (The standardized Iowa Test of Basic Skills is not given until the end of third grade, so the subjects did not have previous-year scores to compare against.)

The results obtained with the pilot version were surprisingly good, according to school-administered pre- and post-tests. Of the 8 pre-tested pilot subjects, one transferred out, and one was not available for post-testing due to behavior problems. The 6 remaining subjects were post-tested in June 1997 by the school reading specialist, using a MacMillan Informal Reading Inventory to measure accuracy, comprehension, and reading rate, as well as phonemic awareness, letter recognition, and letter-sound relationships. After using the Reading Tutor for under eight months, these subjects had advanced by an average of about two years in instructional reading level (defined as

the grade level of material they could read with at least 75% accuracy and 75% comprehension).

These results were consistent with improvements in fluency of assisted reading. In early grade levels, the fluency reflected in such measures is highly correlated with comprehension. To estimate changes in fluency, we analyzed the 6 gigabytes of pilot study data captured by the Reading Tutor (on removable GB JAZ(TM) disks). This data included digitized oral reading by the children, time-aligned output of the speech recognizer, and timestamped event logs of interactions with the Reading Tutor. We estimated students' performance improvements from this data (10498 utterances, 139133 aligned words) as described in [Mostow & Aist, AAAI97]. To control for word difficulty, sentence memorization effects, and word recency effects, we compared a student's first and last encounter of each word, excluding sentences encountered before and words seen earlier on the same day. On average over all eight pilot subjects, 110 days elapsed from the first to last such encounter, with a 16% relative increase in accuracy, and a 35% decrease in inter-word latency (significant per-subject at 95% for 7 subjects, and at 90% for the other).

This was a small-group pilot experiment rather than a large controlled study. Each student used the Reading Tutor for approximately 30-60 sessions, averaging 14 minutes each. This modest amount of interaction seems inadequate to account by itself for the children's dramatic gains in reading. We postulate that the Reading Tutor acted as a catalyst that helped the students gain more from their classroom instruction.

Conclusion

The natural character of the Reading Tutor's spoken dialog contributed to its ability to remain compelling over time for students who had previously found reading an exercise in frustration. Prior to using the Reading Tutor, these children had “tremendous difficulties with traditional reading programs. They now could make meaning out of what they read. They now could have success in the classroom” [Dr. Gayle Griffin, Principal, Fort Pitt Elementary School, in videotaped interview, July 3, 1997].

Acknowledgements

This material is based upon work supported by NSF under Grants IRI-9505156 and CDA-9616546, and by DARPA under Grant F336159311330. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Defense Advanced Research Projects Agency. We thank Fort Pitt Elementary School, our colleagues, and countless others who have helped Project LISTEN.

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

[CAHM 97] G. S. Aist and J. Mostow. A time to be silent and a time to speak: Time-sensitive communicative actions in a reading tutor that listens. In *AAAI Fall Symposium on Communicative Actions in Humans and Machines*. Boston, MA, November, 1997.

[CALL 97] G. S. Aist and J. Mostow. 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, September, 1997.

[Malone 81] T. W. Malone. What makes computer games fun? *BYTE Magazine*, December, 1981, pages 258-277.

[ISGW97 CRLT] J. Mostow. Collaborative Research on Learning Technologies: An Automated Reading Assistant That Listens. *Proceedings of the NSF Interactive Systems Grantees Workshop*, Stevenson, WA, August, 1997. At www.cse.ogi.edu/CSLU/isgw97/reports/mostow3.html.

[ISGW97 IS] J. Mostow. Guiding Spoken Dialogue with Computers by Responding to Prosodic Cues. *Proceedings of the NSF Interactive Systems Grantees Workshop (ISGW97)*, Stevenson, Washington, August, 1997.

[AAAI97] J. Mostow and G. Aist. The Sounds of Silence: Towards Automated Evaluation of Student Learning in a Reading Tutor that Listens. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI-97)*. American Association for Artificial Intelligence, Providence, RI, July, 1997. Pages 355-361.

[CMMII 97] G. S. Aist. Challenges for a mixed initiative spoken dialog system for oral reading tutoring. In *Computational Models for Mixed Initiative Interaction: Working Notes of the AAAI 1997 Spring Symposium.* March, 1997.

[UIST 95] J. Mostow, A. Hauptmann, and S. Roth. Demonstration of a Reading Coach that Listens. In *Proceedings of the Eighth Annual Symposium on User Interface Software and Technology*, pp. 77-78. Sponsored by ACM SIGGRAPH and SIGCHI in cooperation with SIGSOFT, Pittsburgh, PA, November, 1995.

[AAAI 94] J. Mostow, S. Roth, A. G. Hauptmann, and M. Kane, "A Prototype Reading Coach that Listens", *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, American Association for Artificial Intelligence, Seattle, WA, August 1994, pp. 785-792. Recipient of AAAI-94 Outstanding Paper Award.