When Listening Is Not Enough:
Potential Uses of Vision for a Reading Tutor that Listens

John Kominek, Gregory Aist, Jack Mostow

Project LISTEN
215 Cyert Hall, Language Technologies Institute
Carnegie Mellon University
4910 Forbes Avenue
Pittsburgh, PA 15213

jkominek@cs.cmu.edu, aist@cs.cmu.edu, mostow@cs.cmu.edu

Abstract

Speech offers a powerful avenue between user and computer. However, if the user is not speaking, or is speaking to someone else, what is the computer to make of it? Project LISTEN's Reading Tutor is speech-aware software that strives to teach children to read. Because it is useful to know what the child is doing when reading, we are investigating some potential uses of computer vision. By recording and analyzing video of the Tutor in use, we measured the frequency of events that cannot be detected by speech alone. These include how often the child is visually distracted, and how often the teacher or another student provides assistance. This information helps us assess how vision might enhance the effectiveness of the Reading Tutor.

Introduction

Speech-aware software listens to what a person says, and how and when they say it, then takes action based on what was heard. Voice dictation is speech-aware software. When the computer talks back, interaction becomes increasingly social, emulating aspects of human conversation. Over-the-phone systems that supply bus route information produce speech output by necessity; when answers are offered in response to spoken requests (as opposed to key presses), it becomes minimally conversational.

In both of these examples, the goal is transfer of information: first from the user to the computer (e.g. "I want to take the bus to the airport."), then in return ("The 28X will next pass by at 8:45."), with clarification in between ("Where are you located?"). With Project LISTEN's Reading Tutor [Mostow *et al*. 1993, Mostow *et al*. 1994, Mostow *et al*. 1995], our goal is different. What we seek to instill is language knowledge and reading skills. The Reading Tutor listens to children read aloud, and responds with interventions modeled in part after human tutors. With speech awareness central to its design, interaction can be natural, compelling, and effective [Mostow and Aist, PUI 1997].

The Reading Tutor is now deployed in classroom field studies, from kindergarten through fourth grade. Observing real-life usage has allowed us to discover interaction patterns that did not emerge in laboratory settings, or in the more closely monitored pilot study of 1996-1997 [Mostow and Aist, AAAI 1997]. Some interaction problems have been reduced, improving the program's usability. Some technical problems can be addressed by improved speech technology, thus improving the Reading Tutor's timing, and classification of words as having been read correctly, read incorrectly, or omitted.

Other problems are more fundamental – cases where speech input provides insufficient information for an appropriate response. For example, if sound has not been heard for twenty seconds, does this mean that the student is reading silently, is distracted elsewhere, or has left without warning? Vision can discriminate between these cases, but speech alone cannot. Thus even when speech is central to an application, listening may not be enough.

To assess how vision might enhance the effectiveness of the Reading Tutor, we recorded video of students using the Tutor in a classroom setting. In examining the video we concentrated on identifying events and interactions not available from speech.

A Reading Tutor that Listens

The Reading Tutor runs on a stand-alone Windows NT platform. The child uses a headset or handset microphone and has access to a mouse, but not a keyboard. In normal usage the tutor displays a sentence, listens to the child read it, provides help on its own initiative (based on student performance), or in response to student requests – and then proceeds to the next sentence once the child has successfully read the current sentence. To start using the Tutor, the student puts on the headset, selects his or her name, and then picks a story to read. The user can read a word aloud, read a sentence aloud, or read part of a sentence aloud. The user can click on a word for word-specific help, or on the *Help* bubble for assistance on the sentence. Clicking on *Back* moves to the previous sentence; clicking on *Go* moves to the next sentence. The student can click on *Story* to pick a different story, or on *Reader* to end a session.

The Tutor can choose from several communicative actions, involving speech, graphics, and navigation [Aist and Mostow, CALL 1997]. By using a combination of synthetic and digitized human speech, the tutor can read an entire sentence, or provide help on a specific word. Help takes several forms, including letter-by-letter spelling, phonetic decomposition, a rhyming hint, or speaking the full word. Gray shadowing indicates where the Tutor believes the child is in the sentence. In addition, words that have been correctly read are colored. When all of the "important" words have been read correctly (words in a stop list are ignored), the Reading Tutor displays the next sentence. The student can also navigate using the *Back* and *Go* buttons. Backchanneling (e.g. when the tutor whispers "mm-hmm") attempts to elicit response from the student during periods of inactivity [cf. Ward 1996].

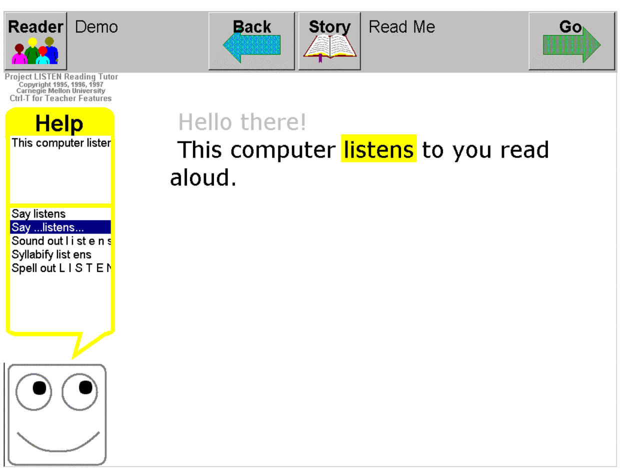

Figure 1. Reading Tutor, October 1997 version.

When Listening Is Not Enough

From the log files and recordings of the Reading Tutor in use, and also from on-site observation, several problems have emerged that are not easily solvable with speech input alone. Some interfere with the sound quality, others with interaction and attention. One observation – double usage – even undermined a critical assumption upon which the software was designed.

Situation	Problem
Microphone too far away from mouth	Volume is faint, impeding speech recognition
Microphone too close to mouth	Sound is clipped, impeding speech recognition
Student touches or grabs the microphone	Sound is corrupted with bursts of noise
Student is chewing gum	Noise confuses the speech recognizer
Student is talking to other people	Reading Tutor assumes student is always talking to it
Student is looking away from the screen and echoing the Tutor	Student is parroting, not reading
Student spins in chair	Student is inattentive; headset cord gets tangled
Student turns around, looks away from Tutor	Student is distracted and inattentive
Two students are using the Tutor at the same time	Who is talking, and who is using the mouse?
Student is not speaking	Student may be thinking, or may be distracted

We assumed that readers would read alone, with occasional brief assistance. Sometimes, however, outside students provide assistance during the entire session. This can be a good situation, where an older, experienced user helps a less experienced user. But consider one extreme case (actually observed on October 24, 1997): one student is wearing the headset while a second is operating the mouse. The second student supplies correct responses to the first, who in turns echoes it to the computer. We call this the "Cyrano Effect" (after the play *Cyrano de Bergerac*). In such circumstances, the tutor's assessment of the student does not accurately reflect their true abilities.

The modalities of input constrain how software can respond. If all that software can handle is mouse clicks, it has a narrow bandwidth from which to infer what the student is doing. If the software can listen, it receives much more information – as long as the student speaks. If the student is not speaking, what is he or she doing? Unlike speech, video input does not rely on the student actively doing anything other than being there. You can stop talking, but you can't turn invisible. Video input may thus allow the Reading Tutor to distinguish between cases where the student is silent and looking at the screen, from cases where the student is silent, but not looking at the screen.

Collecting Video: Three Configurations

Though the Reading Tutor is not now vision-enabled, we are collecting video to help study potential uses of vision. We have used three different recording setups. In the first, a camcorder is placed behind the student, supported by a tripod at an adult's eye view. This setup captures a view similar to a teacher (or researcher) engaged in passive observation. The computer screen is in the camera's field of view. Project LISTEN has used this setup to record users under various conditions: in our computer laboratory, in an elementary school reading room, in minimally occupied school classrooms, and in active classrooms during the school year.

In a second setup, the camcorder is placed behind the computer, facing the user. It is visible to the student. We used this to explore how children would react to being "on camera." Interestingly, students seem to attend to the Reading Tutor more in this situation compared to when the camcorder is behind them and out of sight. Perhaps a camera conspicuously placed behind the student is too much of a distraction. Obviously, a large camcorder on a tripod is not practical for long-term use in a busy classroom.

In a "tutor's eye view" configuration, a fist-sized videoconferencing camera is placed on top of the computer monitor and pointed downward to bring the student's face into central view. A small camera is less conspicuous than a camcorder, and reflects a more normal configuration. So that we could examine the video output during setup and periodic checks, a 12" TV monitor was positioned to the side, oriented away from the reader, and turned off when not in use by a researcher.

Capturing video from the vantage point of the computer serves several purposes.

- *To determine whether the presence of a small camera will adversely affect the student's attention.* Normally, it does not. Students occasionally waved goodbye to the camera, but during reading mostly ignored it. We have found, however, that a TV monitor can be very disruptive – not so much to the student, but to the rest of the class. Turning on the TV even briefly immediately drew an audience eager to see what was going on.
- *To evaluate how students interact with the Tutor when no visitors are watching.* As adult strangers in the classroom, project team members strive to avoid disturbing students working with the Reading Tutor – but our mere presence may have an impact on their behavior. In particular, they look to us for assistance or simply to lodge complaints. We want to construct a picture of how students interact with the Tutor when no visitors are present.
- *To see if video might serve as a substitute for explicit user evaluations.* Elementary students present special challenges for remote evaluation of software. For example, the inability to read or type precludes filling out standard critical-incident reports [Hartson *et al*. CHI 1996]. As an alternative to filling out forms, the Reading Tutor could record video of students saying what they did or did not like, but this option assumes that children have the willingness and ability to articulate their impressions. Another (more difficult) possibility is to monitor facial expressions to trigger automatic video recording.
- *To collect training data for vision algorithms.* Students in a classroom do not sit still, and do not sit squarely in front of the computer. They will look up, look down, turn around, place hand in front of face, pick their nose – and make a vast variety of sounds and facial expressions. In the machine vision community it is standard practice to train algorithms on images of adults, typically graduate students. Our subjects offer challenges outside the range of normal training data.

Potential Uses of Vision

These are some questions that video input can help answer.

- *Is someone using the Tutor at all?* A long period of silence does not imply that no user is currently on. Yet, just because someone is in front of the computer does not mean he or she intends to use the Tutor. The person may simply be sitting down.
- *Who is the user?* Before beginning to read, the Tutor has the child select their name from a list – a task that has proven surprisingly difficult to master. Given highly accurate face recognition, the sign-on procedure could be circumvented.
- *Is the reader interacting with the Tutor, talking to a classmate, or otherwise distracted?* Body position may help reveal the focus of a student's attention.
- *Where is the student looking?* A child will be on-task, most of the time, when looking at the screen or near vicinity. Looking away often indicates off-task behavior, but not always: the child could be listening to instructions on how to use the Tutor.
- *When reading, what word is the student looking at?* Some students will silently pre-read a sentence before reading it aloud. This takes time. Judging by the silence, the Tutor may think the child is stuck.
- *Are other people nearby? Who are they?* Naturally, it is better that the student is speaking to the Tutor rather than chatting to friends. In contrast, if the teacher and student are talking, the Tutor should properly defer.

- *What is the emotional state of the reader?* Boredom, anger, delight, frustration, confusion, etc. – a good human instructor pays attention to cues of these emotions. By design, the Tutor will have the student repeat a sentence until it is read correctly. Yet the computer's judgement of correctness is fallible. If it could see when the child is visibly annoyed, the Tutor might do well to take the hint and move on.

Preliminary Results

Since September 1997, the Reading Tutor has undergone in-classroom evaluation at Fort Pitt Elementary School, Pittsburgh, PA. We installed one computer in each of eight selected classrooms, ranging from kindergarten to grade four. Usage of the Reading Tutor is left to the discretion of the teachers and school principal. (We encourage regular use, but do not impose school policy.)

During November-December 1997 we recorded video through the "tutor's eye view," selecting one of the third grade classrooms for intensive study. By this time, the students in this class had become well acquainted with the Reading Tutor, using it on a regular basis. Session times ranged from 45 seconds (just going through the motions), to 45 minutes (deeply engaged). When done, the reader informs the next student that the computer is free.

From this third grade classroom we have 52 hours of recorded videotape. Students were out of the room for 17h 40m (during lunch, for example). Of the remaining 30h 20m, the tutor was in use for 17h 50m. Such a relatively high rate of usage (59%) was one reason we selected this classroom for study.

In reviewing these sessions, we identified instances of non-standard usage – that is, when the student is not directly looking at the Tutor. A summary of events is tabulated below.

Activity	Occurrences
Number of reading sessions	149
Student glances away (for less than 2 seconds)	602
Student is distracted (for more than 2 seconds)	145
Student adjusts equipment	133
Outside student interferes	79
Student moves partition	21
Student stands up during session	17
Student "monkeys" with equipment	17
Teacher provides assistance	14
Another student offers assistance	8
Teacher removes a student interfering with the reader	2
More than one student using tutor	2

Discussion

We captured a total of 149 reading sessions, with an average duration of just over seven minutes. This includes time spent manipulating equipment at the beginning and end of each session. In the class under study, twenty-six students use the Reading Tutor.

On average, a student is distracted once per session (145/149). Causes include another person approaching nearby, being spoken to directly, loud noises such as the PA system, and movements of students in the background. Many of the events listed as outside student interference (79/149) caused a distraction, but not all; Tutor readers often ignore classmates. Regardless of the cause, a student is considered distracted when their focus has shifted away from the Tutor longer than momentarily. A shift of less than 1-2 seconds is registered as "glancing away." On average, glances occurred four times per session (602/149), or about two times per minute.

As might be expected, individual behavior varies. The focus of some students did not waver once during reading. For others, background activities distracted every twenty to thirty seconds. It is not realistic for the Reading Tutor to expect undivided attention.

Nearly once per session (133/149) the student adjusted a piece of equipment. This may involve moving the keyboard aside, for example, or resetting an egg timer used by the teacher to control session length. Most often the students adjusted the microphone headset worn during reading. (The count excludes initial setup and removal). These ordinary adjustments are in contrast to when the user "monkeys around." We have recordings of students singing into the microphone, and footage of students sticking their tongue at the camera. One bored reader even tried to eat the camera!

Excessive attention to equipment tends to occur after a change has been made to the computer's physical setup. Once students had acclimated to the presence of a monitor-mounted camera, we identified irregularities of usage no more than once per day.

One thing apparent in our tapings is that teachers are busy people. We knew this already, but also knew that our sudden appearance provoked a change of behavior from teachers. When a researcher from Project LISTEN appeared at the door, the teacher checked to see if a student was on the computer. If not, one soon was. Thus, we wanted to learn to what extent a teacher interacts with the Tutor during the normal course of events. Clearly, with just sixteen interventions detected in eighteen hours of use, the Tutor cannot assume (or depend upon) adult assistance.

Given that teachers are usually tending to other tasks, interference from non-reading students becomes a concern. Of 79 events, only two egregious cases provoked corrective action. Not all interruptions proved serious (some children show a remarkable ability for ignoring the pestering interlopers), but users of the Reading Tutor seldom finish a session without experiencing at least interruption.

When used in a classroom setting, the Reading Tutor operates in an active environment. This environment is highly dynamic: people come and go, background noise is high, the visual field is busy. To help isolate the reader from the rest of the class, the third grade teacher placed a moveable vertical partition behind the reader's chair. (The barrier was approximately four feet tall by four feet wide.) In 21 cases the student moved the partition, often when reading a story. Sometimes the reader moved the partition aside, other times closer. Users of the Reading Tutor, evidently, want some control over their environment.

When *Not* to use Vision

One lesson that became painfully clear is that the microphone headsets we employed are ill suited for use by children. The headsets offer six degrees of freedom for adjustment, and are fragile pieces of equipment. At times the headset was on the child's head crooked, or the microphone boom was loose, or the mouthpiece was directly touching lips. We have observed students clutching the microphone during an entire session. Despite adult instruction and correction, few students learned the delicate art of headset adjustment.

For a time, we planned to include "about the headset" lessons as part of the reading material, including video clips showing proper placement. We also considered machine vision as a means for detecting problems, using it to initiate corrective feedback. This is a complicated, heavy-handed approach, and a good example of when *not* to use vision.

The simplest solution is to employ better microphones. We are now evaluating a telephone-like handset.

Some Interesting Events

In the Preliminary Results section, above, an event made it into the table if it constituted something "non-standard." Some of these events must be seen to be fully appreciated.

Nevertheless, consider the following five descriptions from the perspective of: if your favorite speech/vision algorithm saw these sequences, what would it do?

- *Mouthpiece bulb falls off*. A girl is selecting a new story to read, without much apparent interest. She grabs the headset microphone stalk. The sponge mouthpiece bulb falls off, bouncing under the table. She looks down, perplexed, and mutters "hmmm." *Should I go after it?* The Reading Tutor pipes in, "Please say that again." Her attention snaps back. She begins to read: "I have a dream. ...That ALL men are created equal*!" Who cares about the bulb?*
- *Novelty of handset microphone*. We have experimented with three types of microphones: a table mounted mic, a headset, and a handset. Each device carries distinct affordances, and, interestingly, children react to them accordingly. The table microphone approximately mimics a stage microphone. Some children took the cue and began singing, looking around to see who's watching. The accompanying headphones induced one girl to dance around in her seat as if, we hypothesize, swinging to music. And, upon seeing the handset microphone for the first time, one boy tapped imaginary buttons on the inside surface and struck up this conversation.

  > "Hello Dad, may I speak to Mom? All right Mom, I'm on Project LISTEN now. Yeah. Yeah. Uh-huh. Uh-huh. Uh-huh. I don't know where the microphones are though. ... [the student begins reading] *About the Reading Tutor*."
- *Tutor talking to itself*. A student can get up and leave during the middle of a story. The Tutor has no idea that this has happened. Not knowing better, it will try to elicit a response from a student who is no longer there.

  > "... Please say that again. ... I'm waiting for you to select your name or ID. ... To pick your name or ID, move the cursor over it and click the left mouse button. ... Please ask the teacher for help on what to do next." [Tutor then automatically logs out.]

  This sequence of elicitations is quite reasonable (the Tutor is in sign-on mode), except that a new student sat down exactly when the Tutor said to ask for help.
- *Impatience with the Reading Tutor*. One girl obviously wanted to end her session quickly. She was fidgeting in her chair and glancing expectantly at the door. While reading the story "Before I Go to Bed," she

stopped reading to the Tutor and began to talk at the Tutor. Her dialog went like this:

"... *I washed my face. I washed my face*. ... Oh come on man. I'm not trying out much. ... *I washed my face. ... I washed my face*. Oh come on now. It's stuck! Come on, I got to go to the bathroom! Come on! How come it won't go?"

- *Two girls fighting over the Tutor*. The class is returning from lunch and one girl heads to the computer. Another girl cuts her off.

"Hang on, hang on. It's my turn to go on the Project."
"No, you just checked off your name."
"Nuh-uhn."
"No I was getting ready to go on."
"Nuh-uh! Uhm, uhm, uhm. Kiera went to get me and I was getting ready to go on Project LISTEN but I lined up for Science. It's my turn!"

Challenges for an Intelligent Environment

For a computer to be considered intelligent, it should be perceptually aware and be able to respond appropriately in a wide variety of circumstances. Intelligent software must first be able to detect and characterize relevant events, and do so in an environment full of noise.

Some detection tasks are easy. Others are harder, involving difficult distinctions and a complicated sequence of inferences.

- *Student glances away*. This is relatively easy to detect since the action is clear and the duration short (less than 1-2 seconds). Yet there is a difference between glancing away from the screen and looking into the distance, versus glancing at a nearby object on the computer or desk.
- *Student is distracted.* This is almost the complement of glancing away. It is also easy to distinguish, since the duration is longer and the whole body may rotate (rather than just a neck-swivel). However, large changes in head orientation will cause trouble for lip tracking and gaze tracking algorithms, among other things.
- *Student adjusts equipment.* The duration of this event may be long, and the reader may be interacting with another person, but the activity is still related to the task of reading. Unless, that is, it becomes so annoying that the Tutor is completely ignored until the adjustment problem is solved.
- *Student "monkeys" with equipment.* Perhaps the Tutor should respond with "Hey kid, take your hand away! I can't see!" Or, if the camera is knocked out of position and ends up pointing at the ceiling, how should an intelligent agent respond?
- *Outside student interferes* vs. *outside student helps.* Since additional students are often in the periphery, it is difficult to tell what they are doing. The activities *help* versus *interference* depend on subtleties of interaction. Distinguishing on-task from off-task speech is not always easy.
- *Teacher provides assistance.* The teacher plays the most important social role in a classroom, directing class activities. As part of an intelligent environment, the Tutor should be cognizant of social roles and be able to identify who's who.
- *Student stands up during session* or *student moves partition.* These events are readily confused with end-of-session activities. It's not initially obvious that the student is not about to leave.

General Lessons

Beyond the particulars of our application, three major lessons are relevant to the researcher interested in intelligent environments.

First, people are difficult targets. A truly intelligent environment needs to detect expressions, behaviors, interactions and intentions – highly difficult features. Identifying where and who people are is certainly useful, but some of the most interesting events are complex and subtle.

Second, real-world environments are highly active, rich in events and interactions. Simply put, the data is messy. This makes it difficult to distinguish foreground from background events, and thus distinguish what is relevant from what can be ignored.

And third, intelligent environments are dynamic environments. For example, if a person is conversing with a computer and then suddenly yells at someone across the room, the active environment has changed. The second person was once a part of the background, but is now a part of the foreground. The boundaries of an intelligent environment are not fixed.

Conclusion

Speech offers a powerful avenue between user and computer. However, if the user is not speaking, what is the computer to make of it? When an application need only respond to user-initiated action, then patiently waiting is adequate. But if the application is trying to achieve a cooperative goal – here, getting the child to read – then it is useful to know what the user is doing. Listening is not enough when additional information is required for the software to behave an intelligent manner.

The eighteen hours of video that we analyzed helped meet the objectives we posed at the outset of our investigation. We garnered a candid appreciation of how students interact with the tutor when we are not present. Off-task conversations and facial expressions provided dramatic student feedback. We have collected video data for algorithm training and testing. Plus, the data makes us appreciate the steep challenge posed by visual perception operating in a realistic environment.

Our observations are limited in that we intensively studied only one third grade class, and not until the class had become familiar with the Tutor. Now that we have honed our sense of what to look for, we can transport our analysis to lower grades. This will help us reach students who currently find the Reading Tutor difficult to use, yet need its assistance and tutelage the most.

Acknowledgments

This material is based upon work supported by NSF under Grants IRI-9505156 and CDA-9616546, by DARPA under Grant F336159311330, and by the second author's National Science Foundation Graduate Fellowship and Harvey Fellowship. 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 or the official policies, either expressed or implied, of the sponsors or of the United States Government. We thank Fort Pitt Elementary School, our colleagues, and countless others who have helped Project LISTEN. We especially thank Dan Barritt, who ran the Project LISTEN portion of the 1997 Summer Reading Clinic at Fort Pitt Elementary.

References

Aist, G. 1997. Challenges for a Mixed Initiative Spoken Dialog System for Oral Reading Tutoring. *AAAI 1997 Spring Symposium on Computational Models for Mixed Initiative Interaction.* AAAI Technical Report SS-97-04.

Aist, G. S., and Mostow, J. 1997. Adapting Human Tutorial Interventions for a Reading Tutor that Listens: Using Continuous Speech Recognition in Interactive Educational Multimedia. *CALL 97: Theory and Practice of Multimedia in Computer Assisted Language Learning*. Exeter, UK.

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.

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.

Hartson, H. R., Castillo, J. C., Kelso, J., Kamler, J., and Neale, W. C. 1996. Remote evaluation: The network as an extension of the usability laboratory. CHI 96: Vancouver, British Columbia, Canada.

IBM. 1997. IBM Lead Story: Watch Me Read. http://www.ibm.com/Stories/1997/06/voice5.html.

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.

Mostow, J., Hauptmann, A. G., Chase, L. L., and Roth. S. 1993. Towards a Reading Coach that Listens: Automatic Detection of Oral Reading Errors. In *Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI-93)*, 392-397. Washington DC: American Association for Artificial Intelligence.

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.

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.

Mostow, J., and Aist, G. S. 1997. The Sounds of Silence: Towards Automatic Evaluation of Student Learning in a Reading Tutor that Listens. *In Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI 97)*, pp. 355-361.

Mostow, J., and Aist, G. S. 1997. In Workshop on Perceptual User Interfaces (PUI '97), Banff, Alberta, Canada, pp. 87-90.

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.

Russell, M., Brown, C., Skilling, A., Series, R., Wallace, J., Bohnam, B., and Barker, P. 1996. Applications of Automatic Speech Recognition to Speech and Language Development in Young Children. *In Proceedings of the Fourth International Conference on Spoken Language Processing*, Philadelphia PA.

Ward, N. 1996. Using Prosodic Clues to Decide When to Produce Back-channel Utterances. *In Proceedings of the 1996 International Symposium on Spoken Dialogue*, 1728-1731, Philadelphia PA.