Demonstration of a Reading Coach that Listens

Jack Mostow, Alexander G. Hauptmann, and Steven F. Roth
Project LISTEN, 215 Cyert Hall, 4910 Forbes Avenue
Carnegie Mellon University Robotics Institute, Pittsburgh, PA 15213-3890
(412)268-1330
mostow@cs.cmu.edu
http://www.cs.cmu.edu/afs/cs.cmu.edu/user/hcii/www/projectlisten.html

Demonstration of a Reading Coach that Listens

Jack Mostow, Alexander G. Hauptmann, and Steven F. Roth
Project LISTEN, 215 Cyert Hall, 4910 Forbes Avenue
Carnegie Mellon University Robotics Institute, Pittsburgh, PA 15213-3890
(412)268-1330
mostow@cs.cmu.edu
http://www.cs.cmu.edu/afs/cs.cmu.edu/user/hcii/www/projectlisten.html

ABSTRACT
Project LISTEN stands for "Literacy Innovation that Speech Technology ENables." We will demonstrate a prototype automated reading coach that displays text on a screen, listens to a child read it aloud, and helps where needed. We have tested successive prototypes of the coach on several dozen second graders. [1] reports implementation details and evaluation results. Here we summarize its functionality, the issues it raises in human-computer interaction, and how it addresses them. We are redesigning the coach based on our experience, and will demonstrate its successor at UIST '95.

KEYWORDS: Speech interfaces for children, continuous speech recognition, education, children, non-readers

GOALS OF READING COACH
Reading is taught orally in grades 1-3 to help children relate printed English to the spoken language they have already acquired. Unfortunately, a shocking percentage of the nation's children lag behind grade level in reading [2] and grow up functionally illiterate, at an annual productivity cost measured in hundreds of billions of dollars [3]. An automated reading coach could give such children hundreds of hours of individualized attention that teachers and parents cannot. Thus LISTEN's eventual goal is to help children learn to read better over time. The goal of the current coach is to help them read a given text.

The coach is designed to provide a combination of reading and listening, in which the child reads whenever possible, and the coach helps whenever necessary, so as to provide a pleasant, successful reading experience. The coach's assistance, modelled after expert reading teachers, is intended to support word identification, comprehension, and motivation.

To assist word identification, the coach speaks words that the child clicks on, gets stuck on, or misreads.

To assist comprehension, the coach rereads sentences where the child had difficulties. It uses digitized human speech, both for its natural quality and to project a supportive personality.

The coach assists motivation by responding to the child's reading with supportive spoken feedback, and by reducing the frustration that unassisted reading poses for struggling readers.

FUNCTIONS OF SPEECH RECOGNIZER
We adapted CMU's connected speech recognizer to listen to children read. The listening capability required by the coach differs from conventional speech recognition, where the task is to guess what the speaker said. Instead, the coach knows the text the speaker is supposed to read, and must perform the following three related functions to provide the reading assistance described above.

To speak the correct word when the reader gets stuck, the coach must track the reader's position in the known text. This task is complicated by the various speech phenomena characteristic of disfluent readers, including substitution, deletion, repetition, hesitation, sounding out, and other types of insertion.

To provide corrective feedback, the coach must detect reading mistakes. We define mistakes as "important words that the reader failed to speak." We do not treat insertions, repetitions, self-corrections, or hesitations as mistakes.

To know when to respond, the coach must detect the end of the sentence. This is a special case of tracking the reader's position. It is not sufficient to wait for a long pause, because young readers often pause in mid-sentence. Nor is it sufficient simply to wait until the recognizer detects the last word of the sentence, both because the reader may misread it, and because speech recognition errors may cause premature detection.

Thus the reading application places novel demands on the speech recognizer. The coach's listening task is easier than conventional speech recognition in the sense that it just has to detect where the reader is and which words were missed -- it doesn't have to identify what the speaker said instead. But this task is also harder since it involves an infinite "vocabulary" of possible substitutions and insertions, both words and non-words, many of which are highly confusable with the correct text. Moreover, mispronunciations and dialect are not considered reading mistakes, and must therefore be tolerated.

ROBUSTNESS: HOW TOLERATE MISRECOGNITION?
Since speech recognition is less than 100% accurate, we must balance the proportion of reading mistakes the coach detects against the frequency of false alarms -- correct

words misclassified as incorrect. We opted to give the student the benefit of considerable doubt. In the evaluation reported in [1], the coach accepted over 96% of the correctly read words, and detected about half the mistakes flagged by a human coach as likely to affect comprehension. This result was obtained on disfluent reading by children despite the fact that the recognizer had been trained on fluently read speech by adults. To improve accuracy, we are training the recognizer on children's speech.

Fortunately, perfect recognition is not required for the coach to be effective. The coach is designed to behave gracefully even in the face of recognizer errors. For example, it never tells the child that a word was misread, since it may have been read correctly. Conversely, it never tells the child that a word was correct, since it may have been misread. At most, it asks the child to reread a suspect word, and then speaks the correct word. This feedback is intended as either confirmatory or corrective, according to whether the rereading was correct or not.

DIALOGUE: HOW TO MIX INITIATIVE?

Structuring a pedagogically effective spoken dialogue involves a tradeoff among at least the following five competing factors.

Cognitive load: Minimize the cost of making and carrying out decisions. Young readers are often unaware of when they need help, or lack the meta-cognitive skills required to decide to click on a troublesome word. When they do so, positioning the mouse takes them several seconds of effort that distracts from reading. Listening to them read makes it possible to assist them proactively. When the student gets stuck on a word, the coach supplies the word so that the flow of reading can resume. "Getting stuck" is operationalized as timing out without the reader's position advancing in the sentence. The timeout is adjustable, and affects the personality of the coach -- 3 seconds is prompt but aggressive, while 8 seconds is patient but sluggish.

Graceful dialogue: Avoid unnecessary interruption. Human tutors and students use subtle nonverbal cues to guide turn taking and interruption. The coach is designed to reduce the need for such cues. For example, the coach displays the text incrementally, adding one sentence at a time. It waits until the end of the sentence to correct misread words, both to give the child an opportunity for self-correction, and to avoid interrupting. It then presents any feedback on this sentence before displaying the next.

Flexibility: Maximize the student's freedom of behavior. The student is in control while reading the sentence, in the sense that the coach is designed to interrupt in mid-sentence only when the student clicks on a word or gets stuck, in which case it speaks the word and resumes listening. Otherwise it waits until the end of the sentence.

Pedagogical guidance: Achieve pedagogical goals. For example, ensure that the child hears every word read correctly, whether by himself or by the coach. The coach assumes control when responding at the end of the sentence. If no mistakes were detected, it proceeds immediately to the next sentence. If a word was missed, it prompts the student to reread it, and then responds by speaking the correct word, as described before. If the reader had trouble with more than one or two content words, the coach simply rereads the sentence aloud, as an aid to comprehension. Once any end-of-sentence interventions are done, the coach displays the next sentence and resumes listening, thereby returning control to the child.

Predictability: Constrain user options to what the system can interpret. For example, incremental display ensures that the speech recognizer knows which sentence the child is reading.

ACKNOWLEDGEMENTS

We thank Matthew Kane and Leslie Thyberg for their design contributions; Raj Reddy and the rest of the CMU Speech Group for Sphinx-II; Paige Angstadt, Morgan Hankins, and Cindy Neelan for transcription; Lee Ann Kane for her voice; Jim Kocher for video production; CTB Macmillan/McGraw-Hill for permission to use copyrighted reading materials from George Spache's *Diagnostic Reading Scales*; the students and educators at Colfax Elementary School, East Hills Elementary School, Turner School, and Winchester Thurston School for participating in our experiments; CMU's User Studies Lab for the use of facilities acquired through NSF Equipment Grant 9022511; and many friends for their assistance. Maxine Eskenazi helped identify usability problems in the old version and supervise collection of children's oral reading. We thank Brian Milnes for implementing much of the new version of the coach, Chetan Trikha for assisting him, Stephen Reed for his technical contributions, Arshad Tayyeb for recording children's speech, Avram Cheaney for PC support, and Chatham Day Camp and Fort Pitt Elementary School for helping us record there.

This research was supported primarily by the National Science Foundation under Grant Number MDR-9154059 and by the Advanced Research Projects Agency, DoD, through DARPA Order 5167, monitored by the Air Force Avionics Laboratory under contract N00039-85-C-0163. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the sponsors or of the United States Government.

REFERENCES

1. 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), Seattle, WA, August, 1994, pp. 785-792. Recipient of the AAAI-94 Outstanding Paper Award.

2. National Center for Education Statistics. NAEP 1992 Reading Report Card for the Nation and the States: Data from the National and Trial State Assessments. Tech. Rept. Report No. 23-ST06, U.S. Department of Education, Washington, DC, September, 1993.

3. USCD. Closing the Literacy Gap in American Business. United States Commerce Department, 1991.