Challenges for a Mixed Initiative Spoken Dialog System for Oral Reading Tutoring

Gregory S. Aist

Computational Linguistics Program
Carnegie Mellon University
Pittsburgh, PA 15213
aist+@andrew.cmu.edu

Abstract

Deciding when a task is complete and deciding when to intervene and provide assistance are two basic challenges for an intelligent tutoring system. This paper describes these decisions in the context of Project LISTEN, an oral reading tutor that listens to children read aloud and helps them. We present theoretical analysis and experimental results demonstrating that supporting mixed initiative interaction produces better decisions on the task completeness decision than either system-only or user-only initiative. We describe some desired characteristics of a solution to the intervention decision, and specify possible evaluation criteria for such a solution.

Introduction

Intelligent tutoring systems face a wide range of decisions, but two of the most basic are deciding when a task is complete and deciding when to provide assistance. For tasks such as programming or algebra, the decision about task completeness is unambiguous for small examples, if difficult to evaluate for larger problems. When the task involves spoken language performance, as does oral reading, deciding when the student has completed a task such as reading a sentence is more difficult. Errors in speech recognition and the difficulty of unobtrusively measuring comprehension combine to make this decision problematic.

The uncertainty surrounding the evaluation of task completeness complicates the problem of when to provide assistance. The tutor can only approximately judge the correctness of the student's performance. Therefore, it must be able to provide a range of responses that ideally convey the correct information without being irritating to students who actually have completed the task correctly. Providing help-on-demand is useful, but students do not always know when they need assistance (see Mostow and Aist 1997 for a discussion of this).

This paper examines these two decisions in the context of the Reading Tutor being developed by Carnegie Mellon University's Project LISTEN. First, we will review some of the relevant literature. Secondly, we will give an overview of the system that we have used to conduct this research. We then present a theoretical analysis of the task completeness decision and experimental results demonstrating that supporting mixed initiative interaction produces better decisions on this task than either system-only or user-only initiative. We describe some desired characteristics of a solution to the intervention decision. We discuss possible evaluation criteria for such a solution. Since the intervention decision is the subject of work in progress, we do not report final results for this decision here. Finally, we explore potential future research questions.

Related Work

Project LISTEN (Mostow et al. 1994, Mostow et al. 1995, Mostow and Aist 1997) is developing an automated tutor that assists children with oral reading. The reading tutor adapts the Sphinx-II speaker-independent continuous speech recognition system (Huang et al. 1993) to listen to the child read aloud, and provides help when needed. Roughly speaking, the tutor displays a sentence, listens to the child read it, provides help in response to requests or on its own initiative based on student performance, and then displays the next sentence if the child has successfully read the sentence.

Russell et al. (1996) describe a project with similar aims, the Talking and Listening Book project, but they use word spotting techniques to listen for a single word at a time. They also require the child to decide when to move on to the next word (fully user-initiated) or completely reserve that choice to the system (fully system-initiated). Our approach is to use continuous automatic speech recognition to listen to fluent or disfluent readings of entire sentences or individual words. In addition, we

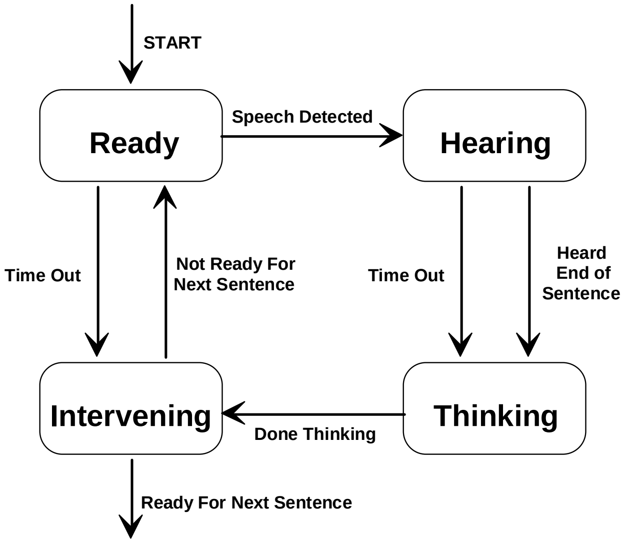

Figure 1. State transition graph for Reading Tutor. For clarity, this graph does not include the effects of help on demand.

allow either the user or the system to take the initiative in moving on to the next sentence.

System Overview

The current version of the Reading Tutor runs on a single stand-alone PC. The child wears a headset microphone and has access to a mouse, but not a keyboard. The system displays a sentence to the child, listens to the child read aloud, and provides spoken and graphical assistance.

The current version of the Reading Tutor has been tested with children in a laboratory setting. The Tutor is currently in place at an inner-city elementary school, where its effectiveness is being assessed in a pilot study of eight third graders who are poor readers. Further usability evaluation is being done with elementary students in the laboratory. Results in this paper are based on data collected in October and November of 1996.

We now describe the input controls available to the user, the internal state of the tutor during interactions with the user, and some of the output behavior of the Reading Tutor.

Input: Controls Available to the User

The system provides a "Remote Control" window with three buttons: *Back* (move to the previous sentence), *Help* (have the tutor read the sentence), and *Go* (move to the next sentence). The user can click on a word for help on it.

Internal State of the Reading Tutor

At any given time during interaction, the tutor is in one of several states. Figure 1 shows the states the tutor can be in during an interaction on a particular sentence. Changes in the shape of the cursor reflect the state of the tutor. For example, the tutor displays an "hourglass" cursor when in the Thinking state.

Output: The Behavior of the Reading Tutor

The Reading Tutor as currently implemented has several options available:

- Reading the sentence
- Reading a word
- Recueing a word by reading the words leading up to it

These interventions employ synchronized audio and visual components, the importance of which is discussed in (Biermann and Long 1996). Other interventions are under development.

Deciding When the Task is Complete

Rather than putting the entire burden of decision on the user or on the system, we allow either one to make the decision to move on. This is a compromise between allowing hands-free use for good readers and providing a learner-centered environment.

User Initiative: Navigational Requests

The user can click the *Go* button on the remote control to move to the next sentence. The user can also click the *Back* button to move back to the previous sentence.

System Initiative: Evaluation of Student Performance on an Individual Sentence

As originally deployed, the Reading Tutor displayed the next sentence when the most recent reading was an acceptable reading of the sentence. For an attempt to be acceptable, all words in the current sentence, except the words on a list of words deemed unimportant for comprehension, must have been read by the student in order during a single utterance. The Reading Tutor gives the student credit for an individual word in the text if it is aligned against an exact match in the output of the speech recognizer using a standard dynamic programming algorithm (Mostow et al. 1993).

The Problem: False Rejection of Correct Attempts

The credit policy in the Reading Tutor as initially deployed resulted in too many false rejections of correct attempts. Students sometimes became frustrated, as we could tell from listening to some of the audio from the pilot study. Teachers also reported the excessive repetitions as a problem. Laboratory observations revealed that there were two causes: continued false rejection of a single word due perhaps to improper lexical modeling of student speech, and false rejection of (seemingly arbitrary) words in long sentences.

The Analysis: Guaranteed Progress in the Limit, Assuming Correct Reading

In order to guarantee continued progress through the story when a child is reading correctly, we conducted an analysis of the internal state of the tutor during the tutorial interaction. First we analyzed the state transition graph, revealing two cycles. Then, we analyzed the behavior of the system in the limit, assuming that the student is reading correctly.

Cycles in the State Transition Graph. Figure 1 shows the state transition graph for the oral reading tutor. Note that there are two cycles:

- Ready → Intervening → Ready

  (e.g. when the student is silent, the tutor prompts the student to read, and then the tutor resumes listening)
- Ready → Hearing → Thinking → Intervening → Ready

  (e.g. when the student attempts to read the sentence, and the tutor does not accept the reading)

These cycles present a threat to the robustness of the system, since students can become stuck in a loop, reading and rereading the same sentence. Russell et al. (1996) dealt with this problem by giving control over moving on to the student. However, their system listens only for a single word at a time. We have observed in tests of our reading tutor that students are willing to read and reread the same sentence repeatedly. Therefore, requiring students to decide when to move on may not scale up from individual words to the continuous oral reading task. In addition, students do not always know when they have or have not read a word correctly.

System Behavior in the Limit. If we assume that the student is reading correctly, and that the speech recognizer has an equal and independent probability of falsely rejecting any given word, we get a straightforward analysis of why the tutor was rejecting student attempts. Since the transition from Intervening to Ready (i.e. rejecting an attempt at a sentence) had a constant probability under the originally deployed tutor, repeated false rejections occurred.

Figure 2 shows the probability of rejecting two consecutive correct attempts with this "local" credit policy, under strong assumptions of independence with respect to the probability of a false rejection both within an utterance and between utterances.

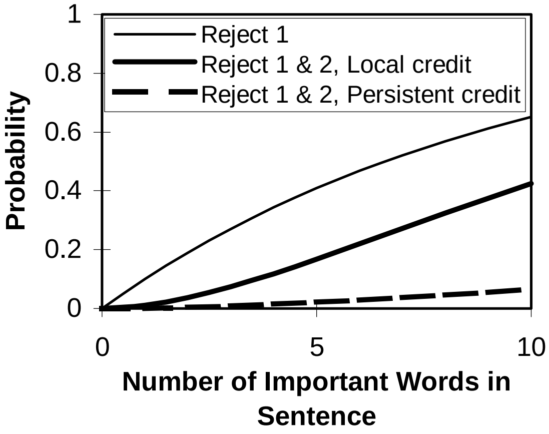

Figure 2. Probability of rejecting two correct utterances in a row, for both local and persistent credit policies. The formulas used to estimate this probability are:

$$P_{\text{local}} = (1- (1-r)^{n})^{k}$$
$$P_{\text{persistent}} = (1- (1-r)^{n})\,(1 - (1-\text{r})^{nr}$$

where n = the number of words in the sentence, k is the number of consecutive attempts, and r is the uniform probability of misrecognizing a word. The equation for persistent credit reflects the fact that, on average, the user will have received credit for all but nr words on a first correct reading. The figure assumes that r is 0.1.

The Solution: Persistent Credit

In order to alleviate student frustration, we devised a mechanism called *persistent credit* that assigns individual credit as described above, but remembers previously assigned credit across attempts. Because the task of reading a sentence is decomposable into subtasks of decoding words, this mechanism provides a natural way of giving partial credit for partially correct attempts. The modified Reading Tutor moves on to the next sentence when the student has received credit for all of the important words in the current sentence, with credit being persistent across attempts. Since the Tutor still gives feedback on the attempt before moving on, it is as accurate as the previous version at detecting children's errors.

Figure 2 shows the probability of becoming stuck on a sentence plotted against the number of words in the sentence. Note that the predicted probability of remaining stuck in a sentence after two correct attempts is substantially lower with the persistent credit policy than with the previous "local" credit policy.

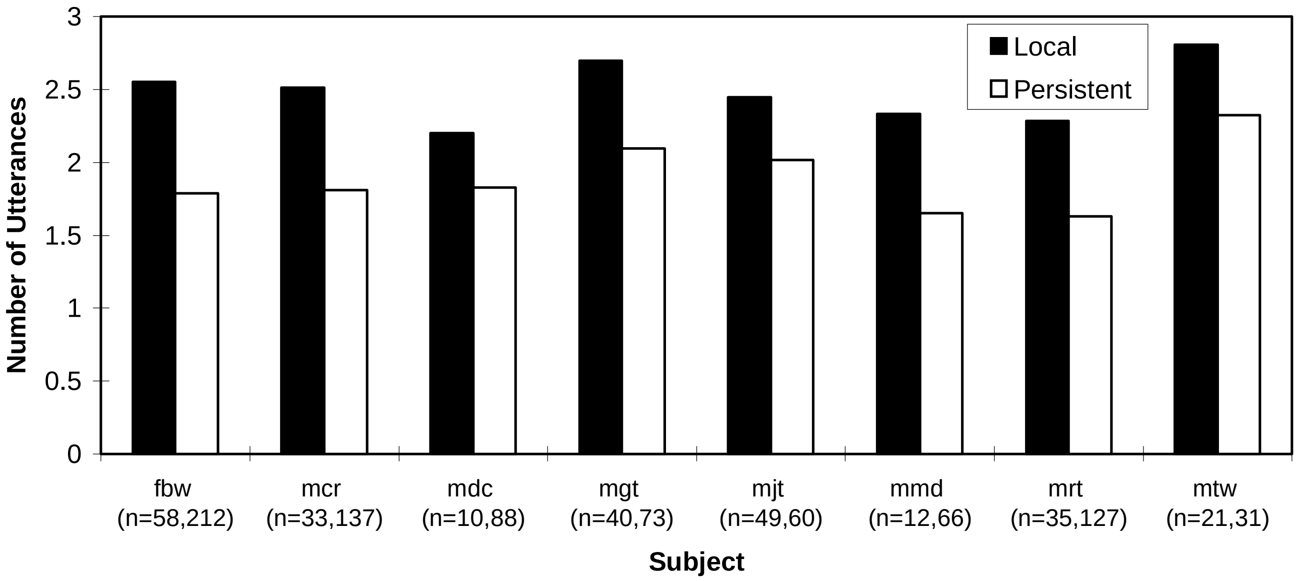

Figure 3. Average number of utterances per sentence, for local and persistent credit, by subject. The number of sentences is shown as *n* in the subject labels. For example, subject fbw read 58 sentences with the local credit version and 212 with the persistent credit version. Differences in *n* reflect different amounts of time spent with the different versions.

Evaluation

We changed the transition behavior of the Intervening state to make the tutor more likely to move on to the next sentence after each student attempt at reading the sentence. Thus, we predict not only a probabilistic guarantee of progress through the instructional task, but in fact fewer attempts per sentence, while maintaining a mixture of system and user interaction. Experimental results support this prediction. Figure 3 shows the average number of utterances per sentence for the Reading Tutor with the local credit policy and with the new persistent credit policy. For every subject, there are fewer utterances per sentence with the persistent credit policy. Observation in the laboratory indicated that user frustration was less evident with the new credit policy.

This solution provides for mixed initiative, unlike Russell et al. (1996). In addition, it reduced user frustration in comparison with the previous local credit model. The Reading Tutor has the motivational advantage of computer-evaluated tasks, and the user-friendliness of a learner-centered system. This is a clear case where a mixed-initiative solution is more desirable than either system-only or user-only initiative.

Deciding When to Intervene

Commercial reading software provides help on demand. We allow the user to ask for help as well as allowing the tutor to provide help on its own initiative.

User Initiative: Help on Demand

The user can click on words for help with the left mouse button. Children frequently use this feature. There is also a *Help* button available that provides help on the entire sentence, but we have not seen children use this feature extensively. The user can also click with the right mouse button to hear what the system heard him or her read for a particular word or for the whole sentence.

System Initiative: Responding to silence

Speech systems that use an open microphone, by necessity, interpret a period of silence at the end of an utterance as the end of a conversational turn. In the oral reading tutoring task, pauses of several seconds in the middle of an attempt at reading a sentence are not uncommon. Therefore, the standard assumptions about turn-taking behavior in spoken language systems do not apply to this task. In addition, there are times when it may be appropriate for the tutor to intervene twice in a row: for example, when the student struggles with a word or is unsure of what to do next. The appropriate length of time to wait during a silence before intervening may depend on several factors, including the student, the difficulty of the text, and the last action taken by the student and the tutor.

The Problem: Limited Conversational Behavior

The Reading Tutor currently makes assumptions similar to those underlying most spoken language systems: essentially, a strict alternation of user and system conversational turns. For example, unlike human tutors, it does not currently provide back-channel feedback (*uh-huh, mmm-hmm*).

However, oral reading is not entirely a turn-taking task. Children begin reading before the tutor stops speaking, and some children read words at the same time the tutor is reading ("choral reading"). As others have suggested (Thompson 1996), a model of dialog will need to account for simultaneous speech by the participants.

The Analysis: Desired Behavior of the Tutor

We intend to redesign the tutor's behavior to allow a more flexible range of conversational behavior. The redesigned tutor should obey the following principles:

- **Be logical.** The tutor should respond to user actions, and initiate actions of its own, in an intuitive and consistent way. Actions should be cooperative (Grice 1975). Complex dialogue interactions will emerge from the actions of the system and the user (Sadek 1996).
- **Be human.** The tutor should actively engage in the learning process, by providing back-channel feedback (Ward 1996) and nonverbal feedback.
- **Be superhuman.** Human teachers often fail to allow sufficient time after asking questions to allow for student response (Stahl 1994). Results from the educational literature (Tobin 1986, Tobin 1987, Gambrell 1983) indicate that allowing wait time of three seconds or more between teacher questions and student responses leads to significant educational benefits. The tutor should be able to pause for appropriate periods ("wait time" (Rowe 1972)) during the dialog to allow for students to think, but should not allow extended, uncomfortable and frustrating silences.

The Solution: A More Flexible Conversational Architecture

This problem is the subject of current research. Results and a more detailed analysis will be reported in (Aist 1997).

Evaluation: Redefining Real-Time Performance

In spoken-language enabled systems, and particularly in intelligent tutoring systems, "real-time performance" implies not simply immediate response, but temporally appropriate behaviors, including adequate pauses between conversational turns depending on the task at hand. Therefore, an evaluation of solutions to the problem of when to intervene will need to include an analysis of how accurate the tutor was in deciding to intervene – specifically if the tutor gave the student enough time to work on the task before offering assistance.

Future Research Questions

Research is underway on several tasks: implementing a richer set of interventions, instrumenting the tutor in order to automatically evaluate usage patterns and dialog flow, implementing an architecture to support more flexible real-time discourse behavior, and collecting speech data for training on children's speech.

Future goals include adapting the behavior of the tutor to an individual student, automatically tracking student improvement in reading ability, and exploring the role of back-channel and nonverbal feedback in human-computer interaction.

Conclusion

Deciding when a task is complete and deciding when to intervene and provide assistance are two basic challenges for an intelligent tutoring system. We have described these decisions in the context of Project LISTEN, an oral reading tutor that listens to children read aloud and helps them. We have presented theoretical analysis and experimental results demonstrating that supporting mixed initiative interaction produces better decisions on the task completeness decision than either system-only or user-only initiative. We have described some desired characteristics of a solution to the intervention decision, and specify possible evaluation criteria for such a solution. Since the intervention decision is the subject of work in progress, we have not reported final results for this decision here.

Acknowledgments

This paper describes work carried out under the supervision of Dr. Jack Mostow (Robotics Institute, Human-Computer Interaction Institute, and Language Technologies Institute) at Carnegie Mellon University. Bryan Nagy and David Sell have contributed programming expertise to this project in order to enable deployment. Jennifer Gutwaks also contributed to research necessary to conduct the pilot study. The Reading Tutor is deployed at Fort Pitt Elementary School in Pittsburgh PA, and we gratefully acknowledge the support and assistance of the Fort Pitt administrators, faculty, and staff. We would like to thank Weekly Reader for permission to use their materials in the Tutor.

This material is based on work supported under a National Science Foundation Graduate 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.

This research was supported in part by the National Science Foundation under Grant Number IRI-9505156, and in part by DARPA under Grant Numbers F33615-93-1-1330 and N00014-93-1-2005. 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.

Finally we would like to thank our subject volunteers, and their parents, for assisting us in a very important way with their participation in the pilot study. We believe that the benefit will be mutual.

References

Aist, G. S. 1997. A General Architecture for a Real-Time Discourse Agent and a Case Study in Oral Reading Tutoring. M.S. thesis, Computational Linguistics Program, Carnegie Mellon University. Forthcoming.

Biermann, A. W., and Long, P. M. 1996. The composition of messages in speech-graphics interactive systems. In Proceedings of the 1996 International Symposium on Spoken Dialogue, Philadelphia PA.

Gambrell, L. B. 1983. The Occurrence of Think-Time During Reading Comprehension Instruction. *Journal of Educational Research* 77(2):77-80.

Grice, H. P. Logic and Conversation. 1975. In A. P. Martinich, ed. *The Philosophy of Language*, 156-167. 3rd edition, 1996. Oxford: Oxford UP.

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.

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. Submitted to the 1997 National Conference on Artificial Intelligence (AAAI 97).

Rowe, M. B. 1972. Wait-time and Rewards as Instructional Variables: Their influence in Language, Logic, and Fate Control. Presented to the National Association for Research in Science Teaching, Chicago IL.

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.

Sadek, M. M., Ferrieux, A., Cozannet, A., Bretier, P, Panaget, F., and Simonin, J. 1996. Effective human-computer cooperative spoken dialogue: The AGS demonstrator. In Proceedings of the 1996 International Symposium on Spoken Dialogue, Philadelphia PA.

Stahl, R. J. 1994. Using ‘Think-time’ and ‘Wait-time’ Skillfully in the Classroom. *ERIC Abstracts*, report number EDO-SO-94-3.

Thompson, H. S. 1996. Why ‘turn-taking’ is the wrong way to analyse dialogue: Empirical and theoretical flaws. In Proceedings of the 1996 International Symposium on Spoken Dialogue, 49-52, Philadelphia PA.

Tobin, K. 1986. Effects of Teacher Wait Time on Discourse Characteristics in Mathematics and Language Arts Classes. *American Educational Research Journal* 23(2):191-200.

Tobin, K. 1987. The Role of Wait Time in Higher Cognitive Level Learning. *Review of Educational Research* 57(1):69-95.

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.