Eindhoven, The Netherlands, 1999.

MEASURING THE EFFECTS OF BACKCHANNELING IN COMPUTERIZED ORAL READING TUTORING

Greg Aist and Jack Mostow
aist@cs.cmu.edu, mostow@cs.cmu.edu

Project LISTEN
213 Cyert Hall
Carnegie Mellon University
Pittsburgh PA 15213

ABSTRACT

What is the effect of backchanneling on human-computer dialog, and how should such effects be measured? We present experiments designed to evaluate the immediate effects of backchanneling on computer-assisted oral reading tutoring. These experiments are implemented in a reading tutor that listens to children read aloud, and helps them learn to read. As a byproduct of designing, conducting, and evaluating these experiments, we are able to describe some unique methodological challenges in evaluating the effects of low-level turn taking dialog behavior.

1. WHY MIGHT SPOKEN DIALOG SYSTEMS BACKCHANNEL?

Incorporating duration-sensitive turn-taking actions, such as prompting, interrupting, and backchanneling, is a critical challenge for spoken dialog systems. Backchanneling in particular might lend naturalness to dialog: a system might provide a sense of presence by saying "mm-hmm" from time to time. In a tutoring system, backchanneling could also provide subtle encouragement to the student. Researchers have investigated a number of methods for generating backchanneling and other turn-taking utterances (Tsukahara ICSLP 1998, Ward ICSLP 1996; cf. Donaldson 1997).

2. PROJECT LISTEN's READING TUTOR LISTENS TO KIDS READ ALOUD

Project LISTEN's Reading Tutor [Mostow and Aist CALICO 1999, Mostow and Aist AAAI 1997] helps children learn to read by providing assisted practice in reading connected text. As figure 1 illustrates, the Reading Tutor presents one sentence at a time, graying out earlier text. Using continuous, open-mike speech recognition, the Reading Tutor visibly shadows the word it expects to hear next, and turns words green that it accepts as read correctly. Its responses are adapted from expert reading teachers, combining recorded human voices with graphical effects such as highlighting each text word as the Tutor speaks it [Aist and Mostow CALL 1997].

A simple persona gives the child a visible audience. To appear animate, it blinks now and then. To appear attentive, the persona gazes at the expected word, the current mouse position, or the student.

If the student pauses before reading the sentence, the Reading Tutor may prompt the student to read, or the Reading Tutor may politely cough to gently remind the student to start reading. The Reading Tutor responds to oral reading by (occasionally) interrupting a student's non-self-corrected mistake [Aist ICSLP 1998], (occasionally) backchanneling to encourage the child after a brief silence, or giving a hint when the child gets stuck. After the student finishes an attempt at a sentence, the Reading Tutor may (occasionally) praise good or improved performance, correct a mistake on a content word, or read the sentence aloud to the child after multiple mistakes.

The student can click on a word for help. The student can also click on the persona's face for help on the sentence. Word help may include saying the word, reading the words leading up to it, sounding or spelling it out, splitting it (visibly and audibly) into syllables, giving a rhyming hint, displaying a picture, or playing a sound effect. Sentence help includes reading the sentence (fluently or word by word) or echoing the student's most recent utterance, for example when the student is narrating a typed-in story.

Navigation buttons increase learner control; the child can:

- Click on Goodbye to log out from the Tutor.
- Click on Back to go back to the previous sentence.

Figure 1. Reading Tutor with screen used for reading in Spring 1998. The summer 1997 version was similar.

- Click on Go to go on to the next sentence of the story.
- (Pre-Summer 1999 versions only) Click Story to choose a new story.

Other clicks elicit suggestions aimed at likely user intent, rather than null responses that might frustrate or confuse:

- Clicking between words suggests clicking on a word.
- Clicking above the sentence suggests the Back button.
- Right- or middle-click suggests the left mouse button.

Prior to summer 1999, researchers or teachers enrolled students. (Summer 1999 and later versions enable students to enroll themselves.) During classroom use, students log on and read one or more stories per session, with session lengths ranging from a few minutes to over half an hour.

3. BACKCHANNELING IN THE READING TUTOR

While the student is reading the story, the Reading Tutor runs a turn-taking procedure five times per second. The turn-taking procedure classifies all situations into sixteen classes, based on four binary variables:

- Is the student taking a turn?
- Is the student speaking now?
- Is the Reading Tutor taking a turn?
- Is the Reading Tutor speaking now?

For example, if the student is reading a sentence out loud and hesitates very briefly, the turn-taking state would be “student taking turn, student NOT speaking now, Reading Tutor NOT taking turn, Reading Tutor NOT speaking now.”

Each of the four turn-taking states is associated with an ordered list of rules. The first applicable rule in the list fires.

The Reading Tutor considers backchanneling according to the following turn-taking rule: when a 2-second student pause occurs while the student is reading aloud, and the student does not appear to be at the end of a sentence, and the Tutor hasn't just backchanneled, consider backchanneling. This rule appears in the rule list for the “student NOT taking turn, student NOT speaking now, Reading Tutor NOT taking turn, Reading Tutor NOT speaking now” turn-taking state. Note that the preconditions of this rule include both prosodic features such as pause duration and task-specific features such as the estimated position of the student in the sentence. The backchannel actions of the Reading Tutor include playing recorded phrases such as "uh-huh" and "mm-hmm". The turn-taking rule for backchanneling coexists with a number of other turn-taking rules, including rules for corrective feedback, interruption, and prompting [Aist ICSLP 1998]. The intent of including backchanneling in the dialog was to encourage the student to continue reading.

4. MEASURING THE EFFECTS OF BACKCHANNELING

What happens after the Reading Tutor backchannels? We knew that several types of events might occur:

- The student may speak next, reading part or all of the sentence with varying degrees of success.
- The student may click on something relevant, such as a word or button.
- The student may continue to hesitate long enough for the Reading Tutor to take a turn.

We would like to know how often these events happen. Furthermore, we also want to know what *would* have happened if the Tutor had not backchanneled -- in order to test for local effects of backchanneling on the dialog.

4.1. Choosing a Control Condition

How does backchanneling affect the dialog, as compared to other possible alternative actions? Evaluating the effects of turn-taking acts is tricky. We have been developing a novel experimental paradigm called "invisible experiments", where a spoken dialog system chooses among several alternative dialog actions and records the effects of its behavior for later analysis [Mostow and Aist AAAI 1997]. With invisible experiments, the choice of control conditions is not always clear. When choosing a control condition for backchanneling in particular, there is no "do nothing" choice available. (Just waiting and saying nothing when the user believes it is the system's turn may itself have meaning.) In addition, it is important to consider how much time the student has to respond in various possible control conditions. We must also consider whether to include the time that the Reading Tutor would have used when backchanneling is included in the total response time allowed to the student.

We summarize the backchanneling behavior of the Reading Tutor as follows:

- Backchanneling. If 2 seconds have gone by after silence is detected, and the Reading Tutor hasn’t just backchanneled, play some backchannel response, such as “uh-huh” or “mm-hmm”. (Playing a sound file for one of these responses typically takes 0.5 - 1 second). Wait 4 seconds for the student to respond, before giving more help (e.g. reading the sentence.)

We have identified several possible alternative control conditions for backchanneling.

- Alternative 1: “Reconsider backchanneling.” Sometimes, don't backchannel. Say nothing and give the student 2 more seconds to respond. Then, consider backchanneling again.
- Alternative 2: “Extra time.” Don't backchannel, but don’t consider backchanneling again. Say nothing and give the student 2 more seconds to respond. Then, take a turn.

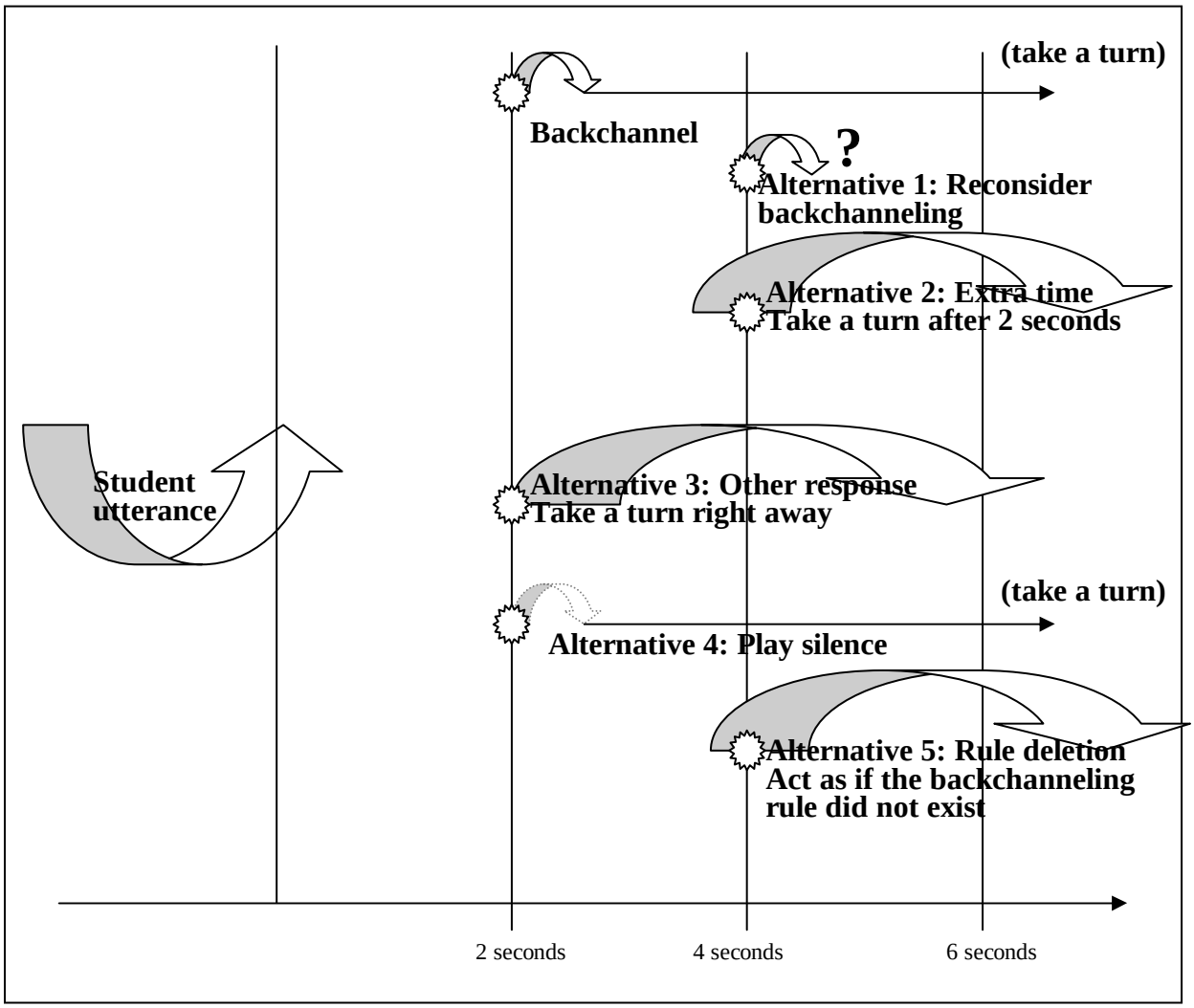

Figure 2. Alternatives to backchanneling.

- Alternative 3: "Other response." Don't backchannel, but take a turn right away instead. (I.e. instead of backchanneling, give corrective feedback.) The rationale here is that if it's time to speak, the system should say something -- either backchanneling or providing corrective feedback.
- Alternative 4: "Play silence". Behave just as if backchanneling, except "play" a silence of the same duration as an audible backchanneling response, and then wait 4 more seconds before taking a turn. This alternative, while seeming artificial, allows the student the same total amount of response time as if the Reading Tutor were playing an audible backchannel.
- Alternative 5: "Rule deletion". Behave as if the Reading Tutor had no backchanneling rule; just follow whatever rule fires next. For example, take a turn after 2 more seconds of silence (for a total pause of 4 seconds.) The rationale here is that *when* to speak may depend on *what* the system is going to say.

In Figure 2, "Extra time" (Alternative 2) and "Rule deletion" (Alternative 5) reduce to the same behavior. The amount of time before the backchanneling rule would fire (2 seconds) plus the 2 seconds of extra time to give in "Extra time" (Alternative 2) is equal to the 4-second turn-taking wait in "Rule deletion" (Alternative 5).

4.2. Backchanneling vs. Other Response

During a reading clinic held in the summer of 1997, we conducted an invisible experiment intended to compare the effects of backchanneling vs. "Other response" (Alternative 3). We evaluated the effectiveness of backchanneling by measuring the relative frequency of what happened next within a 4-second window after a random decision to backchannel, compared to what happened next otherwise, that is, after a random decision to proceed without backchanneling. Our automated analysis of this experiment

was based on 12071 random trials (from one month of use by over 60 students) that occurred when the Reading Tutor's backchanneling rule fired. In 6115 cases the Tutor randomly decided to backchannel and in 5956 cases it randomly decided to proceed without backchanneling. About half the time (47.2% after backchanneling vs. 52.9% after not backchanneling) the student responded within 4 seconds.

This result led us to suspect that when it comes to eliciting student speech, backchanneling is almost as effective as proceeding with whatever the Reading Tutor would do if backchanneling were not an option. But since backchanneling has much lower cost in time, disruption, attention, and robustness, it may be preferable overall. When it fails to elicit speech, the Reading Tutor can still take a turn. Consequently backchanneling seems worthwhile to try, since it doesn't cost much and it often succeeds.

Unfortunately, we discovered a flaw in the turn-taking implementation after the experiment was over. We suspect that if the student utterance was less than two seconds, the Reading Tutor waited a little bit longer before taking a turn due to a software bug. We learned from this disappointment that almost unnoticeable software bugs could lead to inaccurate results if all of the analysis was done automatically. (In fact we discovered the bug only by carefully reading the Reading Tutor logs and actually listening to student utterances.)

On reflection, also, "Other response" seemed like an unfair control condition for backchanneling because having the Reading Tutor take a turn right away did not let us find out what the student would have done if given extra time. We decided that "Rule deletion" (Alternative 5), would provide a comparison that better answered the question "What would have happened if the Reading Tutor had not backchanneled?"

4.3. Backchanneling vs. Rule Deletion

We modified the Reading Tutor to use "Rule deletion" as a control condition, and conducted this new experiment in the spring of 1998. Upon listening to some of the data from the previous study, we decided that various factors might have adversely affected either the generation of the actual backchanneling phenomena we were interested in, or our automated analysis of the resulting dialog. These factors include software bugs, system slowdown, sound from the earphone being recorded by the microphone, excess background noise or background speech being interpreted as a student utterance, and unanticipated post-backchanneling events such as crashes. We decided therefore to conduct a manual exploratory analysis of a randomly selected portion of the backchanneling events. This analysis consisted of examining the log files before, during, and after the backchanneling event, and listening to the sound recorded before and after the backchanneling.

How were backchanneling examples selected? We chose to look at second grade and fifth grade for variety in student age and ability level. We chose classrooms that had identical Reading Tutor configurations in terms of system speed (200 MHz Pentium Pro) and main memory (128 MB RAM). We chose a time period, February 24 through March 31, 1998, that consisted of data from the same December 17, 1997 version of the Reading Tutor (for consistency) collected onto recordable CDs at the same time (for convenience). We then chose days at random, to avoid a length bias towards days with more occurrences of backchanneling because of any (unsuspected) factors related to the day of the week. We skipped days with no Reading Tutor usage or no occurrences of backchanneling. Finally, we chose the first three students who used the Reading Tutor in the set of previously selected days and looked at several examples of backchanneling events from dialogs with each student. All students using the Reading Tutor were supposed to receive equal time on the computer in these classrooms, and this constraint should help reduce any bias towards looking at data only from more frequent users.

What types of events occurred before and after the Reading Tutor backchanneled? The first three instances of Reading Tutor backchanneling from each student in the exploratory data analysis are shown in Table 1. Note that many of the occurrences of backchanneling are either infelicitous due to software bugs (10 out of 18) as when the Reading Tutor backchanneled after taking a turn itself), or due to the Reading Tutor hearing itself (5 out of 18). Only three instances (15, 16, 17) out of eighteen (16.7%) seem appropriate because the student may have paused in the middle of reading a sentence. Evidently the backchanneling rule should be more accurate before further analysis is meaningful.

#	*Student*	*What happened before*	*Backchannel*	*What happened after*	*Event time (1998)*
1	MARW grade 2	RT says, “I didn’t hear you read that word.”	Um-hmm	S clicks for help on sentence	Mar 06 08:52:33.984
2	MARW	RT sounds out ‘Jack’	Um-hmm	S clicks for help	Mar 06 08:52:52.046
3	MARW	RT gives onset & rime for ‘Jack’	(none)	S clicks for help	Mar 06 08:52:56.906
4	MOH grade 2	S clicks Go button. RT says, “Try clicking Help instead.”	Mmm	RT says “B here makes the sound /b/”	Feb 26 08:23:47.109
5	MOH	RT hears itself say, “Makes the sound”	Mmm	RT says “B here makes the sound /b/”	Feb 26 08:23:57.437
6	MOH	RT gives onset & rime for ‘Bob’	Mmm	S clicks dimmed Go button; RT says, “Try clicking Help instead.”	Feb 26 08:24:14.718
7	FKVB grade 2	RT says, “One starts like Ones”	(none)	S clicks dimmed Go button; RT says, “Try reading the sentence instead.”	Feb 26 10:20:16.953
8	FKVB	RT says, “One starts like Ones”	Um-hmm	S clicks on word for help	Feb 26 10:28:05.000
9	FKVB	RT says, “two rhymes with do”	Um-hmm	RT prompts with a cough	Feb 26 10:28:13.718
10	FCR grade 5	RT hears itself say “Might be a little... <sound out TOMORROW>”	Uh-huh	RT sounds out EVEN	Mar 24 10:30:00.890
11	FCR	RT hears itself say “Tomorrow”	Mmm	RT says “Read this sentence with me” and then reads the sentence	Mar 24 10:30:11.265
12	FCR	RT hears itself sound out EVEN	Uh-huh	RT says, “This says…” and then reads the sentence	Mar 24 10:30:26.078
13	FAW grade 5	S clicks middle mouse button	Uh-huh	S clicks dimmed Go button; RT says, “Try clicking on a word instead.”	Mar 10 11:40:02.000
14	FAW	RT hears itself say “/ch/. See”	Uh-huh	S reads sentence: “There was a law in the city of the... Whatever... which gave to its citizens the power of impelling their daughters to marry”	Mar 10 11:40:02.000
15	FAW	S reads sentence: “There was a law in the city of the... Whatever... which gave to its citizens the power of impelling their daughters to marry” (same as previous utterance.)	Uh-huh	S reads part of sentence with a slight pause in the speech: “so ever … they pleased.”	Mar 10 11:40:22.171
16	FCH grade 5	S reads part of sentence: “Silly Pilly go… hold on”	Um-hmm	RT reads sentence: “Silly Pilly goes to school, by Jack Mostow.”	Mar 10 11:57:08.421
17	FCH	S reads sentence: “Oh that's just the said school said Pilly”	Um-hmm	S clicks Go to move on	Mar 10 12:06:14.640
18	FCH	S opens new story	Um-hmm	S reads sentence: “A bed”	Mar 10 12:07:25.546

Table 1. Exploratory data analysis of Reading Tutor (RT) backchanneling and student (S) behavior.

5. CONCLUSIONS

In this paper, we have briefly described the implementation of backchanneling in a Reading Tutor that listens to children read aloud, and helps them learn to read. We have discussed several alternative behaviors and their use as control conditions for backchanneling. Finally, we have described two experiments designed to measure the immediate effects of backchanneling on the reading tutoring dialog.

5. REFERENCES

1. Aist, G. 1998. Expanding A Time-Sensitive Conversational Architecture For Turn-Taking To Handle Content-Driven Interruption. *International Conference on Speech and Language Processing* (ICSLP98). Sydney, Australia, December, 1998.

2. Aist, G. and Mostow, J. 1997. 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.

3. Donaldson, T., and Cohen, R. 1997. A Constraint Satisfaction Framework for Managing Mixed-Initiative Discourse. In *Computational Models for Mixed Initiative Interaction: Working Notes of the AAAI 1997 Spring Symposium*. March, 1997.

4. Mostow, J. and Aist, G. 1999. Giving Help and Praise in a Reading Tutor with Imperfect Listening – Because Automated Speech Recognition Means Never Being Able to Say You're Certain. *CALICO Journal 16*:3, 407-424. Special issue (M. Holland, Ed.), *Tutors that Listen: Speech recognition for Language Learning*, 1999.

5. Mostow, J., and Aist, G. 1997. 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.

6. Tsukahara, W. 1998. An Algorithm for Choosing Japanese Acknowledgments Using Prosodic Cues and Context. In *Proc. 5th International Conference on Spoken Language Processing*, p.p. 691-694. Sydney, Australia, December, 1998.

7. Ward, N. 1996. Using Prosodic Clues to Decide When to Produce Back-channel Utterances. In *Proc. 4th International Conference on Spoken Language Processing* (ICSLP-96). Philadelphia, PA, USA, October 1996.