



A Scientific Learning Whitepaper

300 Frank H. Ogawa Plaza, Ste. 600
Oakland, CA 94612
888-358-0212
www.SciLEARN.com

Reading Assistant: Technology for Guided Oral Reading

Valerie Beattie, Ph.D.
Director of Speech Engineering

Sylvain Chevalier, Ph.D.
Director of Engineering

April 10, 2012
Revised March 13, 2015

The Reading Assistant Advantage

Scientific Learning's Reading Assistant takes a proven technique for improving reading fluency – guided oral reading with feedback – and makes it practical to give a classroom or a school full of students this essential practice, as often as they need it. At the same time, Reading Assistant monitors and summarizes students' progress for teachers and administrators. This key feature would not be possible without Reading Assistant's Guided Oral Reading technology components, including Reading Assistant Speech Recognition and Speech Processing, and Intelligent Reading Feedback.

Reading Assistant Speech Recognition and Speech Processing

Guiding and monitoring a student's reading in real time requires speech recognition software running on the user's computer, and the ability to access and configure the user's microphone. At the same time, a web-based interface allows users to access Reading Assistant through a browser and get new functionality and content easily.

To achieve both of these goals, Reading Assistant uses a lightweight browser plug-in on most browsers to provide speech recognition and related functionality. Starting in March 2015 on the Chrome browser only, speech recognition and related functionality will be provided entirely through web-based technologies, eliminating the need for a plug-in on this browser. This approach may be expanded to other browsers over time as the needed technology features become fully available in other browsers.

Both of these technology approaches use the same Reading Assistant Speech Recognition and Speech Processing components, which are described in the following sections.

PocketSphinx Speech Recognizer

The most important component of Reading Assistant Speech Recognition and Speech Processing is the speech recognition software, which allows Reading Assistant to listen and follow along as the student reads. Reading Assistant uses the PocketSphinx speech recognizer, part of the CMU Sphinx Open Source Toolkit for Speech Recognition (<http://cmusphinx.sourceforge.net/>). The PocketSphinx version is kept as up-to-date as possible to take advantage of the latest speech recognition features and enhancements.

Acoustic Model

A critical component of the speech recognizer is the acoustic model, which represents sounds and sequences of sounds. The acoustic model is created using recorded speech from many different individuals. During recognition, the student's speech is compared to the model to determine what words are being spoken.

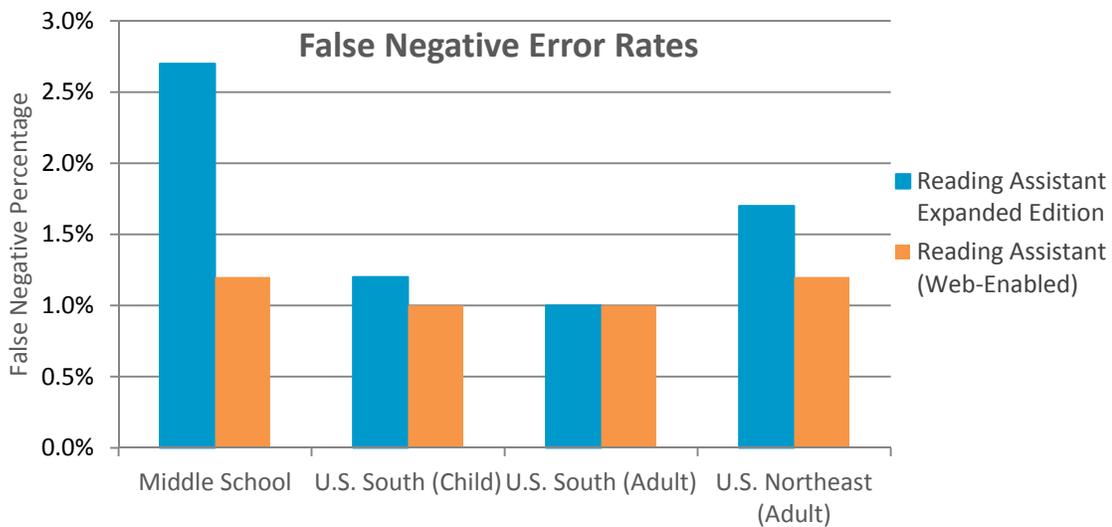
Unlike many other speech recognition applications, Reading Assistant needs to be able to interpret speech from students ranging from five years old to adults. The software also has to handle a wide range of dialects, English language learners, and challenging noise conditions in labs full of students.

To create an acoustic model that could handle all of this variability, a large quantity and broad range of audio data is needed. The table below describes the categories and quantities of data that were used to create a new acoustic model for Reading Assistant.

| Type of Data | Speakers | Number of Hours of Recordings |
|---|--|-------------------------------|
| External Corpora (Several Sources) | Includes adult and child data as well as native and non-native speakers of English | 104.0 |
| Proprietary Data Collection | Southern U.S. Speakers: approx. 80% child or teen ages 6 through 17, 20% adult | 60.2 |

The Southern U.S. data used was from the comprehensive in-house data collection described in “Reading Verification Improvements in Scientific Learning Reading Assistant Expanded Edition” (Scientific Learning: Research Reports, 13(13):1-18). The models contain over 80 hours of data from children and teens. In total, over 160,000 audio files were used in the development of the acoustic model for Reading Assistant.

With the latest speech recognition technology and the comprehensive data used to develop the acoustic model, the web-enabled version of Reading Assistant achieves false negative error rates equivalent to, or lower than, the previous version (Reading Assistant Expanded Edition). False negatives occur when the software gives feedback (intervenes on) a word during guided oral reading, when in fact the word was read correctly. It is critical for this type of error to occur very infrequently, otherwise fluency will be disrupted and frustration will occur. Generally, false negative rates of around one percent or lower are desirable for usability and promotion of fluency in students. Comparison results for four test sets are shown in the table below.



The Middle School test set was comprised of 6th through 8th graders for a total of 586 test audio files (each audio file is generally about a sentence of reading). The test data was collected in the Boston area and had a high percentage of ELL students among the test subjects. The U.S. South (Child) and U.S. South (Adult) test sets were collected as part of the in-house data collection in the Southern U.S. described above and contained 6,876 and 4,762 audio files, respectively. The U.S. Northeast (Adult) test set was from a different in-house data collection in the Boston area and had 1,874 test audio files.

As an added benefit, with the coverage provided by this new acoustic model, students are no longer required to go through a voice enrollment or customization process before recording readings. The speech recognition software is ready to use after just a quick check of the microphone.

Speech Recognition Grammars

Another important component in the configuration of the speech recognizer is the recognition grammars. These specify what words we expect the user to say, and in what sequence we expect them to be spoken. In the case of the guided oral reading task, the software expects the user to be reading the displayed text, but must also allow for the fact that the user may make errors, repeat words, or restart a sentence. The speech recognition grammars allow the software to represent and interpret all of these behaviors, and even assign different probabilities to different ways that the words in the text could be spoken. For example, if a student just read the first word of a sentence in the text, they are most likely to continue with the next word; it's less likely, but also possible, that they will repeat that first word; on the other hand it is pretty unlikely that they will skip several words and read the last word of the sentence next.

It's also possible that users will get distracted and say words that aren't in the story at all. So-called "filler" words are models which represent general speech sounds and are included in the grammars to accommodate this type of behavior.

Cross-Platform Development

The Reading Assistant Plug-in uses the FireBreath framework for creating browser plug-ins. The FireBreath architecture allows the core functionality of the plug-in to be deployed cross-platform (Mac and Windows) and on different browsers. On the Chrome browser, speech recognition and speech processing functionality are implemented within the web application itself, so a plug-in is not required.

Other Speech Processing

Other important speech processing technologies are part of Reading Assistant, including speech compression and automatic gain control.

Speech Compression

Reading Assistant records the student's reading for later playback. In order to be able to upload and save a student's reading, speech compression is used to reduce bandwidth and storage requirements. Reading Assistant uses the Speex encoder, which was designed for Voice over IP (VoIP) applications and which can be encapsulated in a format for browser-based playback.

Automatic Gain Control

Controlling the input microphone gain level (volume) is essential for ensuring good audio quality and a successful guided oral reading experience. If the level is too low, the user cannot be heard; if it is too high, distortion will occur and the student's speech will be unintelligible. Reading Assistant constantly monitors and adjusts the gain level as needed.

Intelligent Feedback During Reading

For a successful guided oral reading experience, the application logic which determines when and how to give feedback is just as important as the strengths of the speech recognition engine.

Striking a Balance: Promoting Fluency, but Intervening When Needed

Reading Assistant must be able to successfully guide students who have a wide range of reading abilities and approaches. The software intervenes (indicates that the user should repeat a word, or provides the word) when needed; however care must be taken to avoid intervening when help is not really needed, as this will interfere with fluency and cause frustration. The details of how feedback is given, and when the software decides to give feedback vs. when the user is allowed to continue, are critical for a positive user experience.

Automatically Adjusting to User and Reading Level

In order to give help appropriately for each student, Reading Assistant adjusts feedback based on the reading level of the content and the student.

Reading Level-Dependent Timing of Feedback

Beginning readers need more time to decode words and apply word attack strategies. On the other hand, more fluent readers who need help on a word will benefit by getting that feedback faster. Thus, Reading Assistant adjusts the timing of feedback depending on the reading level of the content. For reading levels that are at 2nd grade or below, students will generally get three seconds before feedback is given, whereas at higher levels they will get only two seconds.

Reading Level-Dependent Processing of Readings

In Reading Assistant, words in a story are given a category assignment which influences how they are processed and treated in Record My Reading. This assignment can affect logic and settings in the speech recognition software and in the associated processing and feedback modules. “Glue” words are a category typically composed of short, very common words such as articles and prepositions, which it is assumed that the reader already knows. These words are typically not important to meaning, and they are often elided or de-emphasized in speech which results in their being misrecognized. For these reasons, we do not require correct recognition of them to allow the user to proceed in the text.

The list of words generally included in the glue word category varies by reading level, with shorter lists for the 1st and 2nd grade reading levels, and a longer list used for 3rd grade and higher. Shorter lists are more appropriate for readers at the lower grade levels, since only a small number of very common words are likely to be automatic for readers at these levels.

Feedback Timing Adjusts Automatically Depending on User Reading Behavior

In addition to being dependent on the reading level as described in this section, the timing of feedback is also automatically adjusted based on the user’s reading behavior. If the software detects that a user has continued on past an error, it will intervene immediately rather than waiting for the time periods described above (two or three seconds) to elapse. This is called an “accelerated” intervention and keeps more fluent readers from getting too far ahead in the text before we intervene. On the other hand, if we detect that a user may still be trying to sound out or struggle with a word, we will give the user additional time to complete working on the word. This allows users a little additional time to apply word attack strategies. The decision to allow more time is based on whether we detect speech or silence in the audio at the moment when we would give an intervention.

Customizable

If needed, the software can be further customized for each student. The intervention wait time (time before feedback is given) can be adjusted to be either longer or shorter than the default time for the given reading level. In addition, how strict the software is on enforcing correct reading pronunciation can also be adjusted to be more or less strict than the default value.

The innovations in intelligent feedback which are described here were originally developed for Reading Assistant Expanded Edition, and have been carried forward to the web-based version of Reading Assistant. For more details on these important features, please see “Reading Verification Improvements in Scientific Learning Reading Assistant Expanded Edition” (Scientific Learning: Research Reports, 13(13):1-18).

Patented Approach

Scientific Learning’s unique methods for guided oral reading have several patents associated with them, both awarded and pending:

- “Assessing fluency based on elapsed time”, U.S. Patent No. 7433819
- “Intelligent Tutoring Feedback”, U.S. Patent No. 8109765
- “Word competition models in voice recognition”, U.S. Patent No. 7624013
- “Microphone setup and testing in voice recognition software”, U.S. Patent No. 7243068
- “Sentence Level Analysis”, U.S. Patent Application No. 20060069558