Sam Krasnik
2/1/04
cs674

Critique: A Prototype Reading Coach that Listens

One of the first elements of the evaluation that catches the attention of the reader is the selection of test subjects. The authors of the reading coach selected a group of second graders to test their system on. While many of the possible troubles with the graphical interface to the reading coach will become more quickly apparent with younger children, there are other drawbacks that outweigh the positive aspects of the selection.

Most importantly, it is not obvious as to how the system will scale when used with readers in higher grades. Firstly, the vocabulary used in 2nd grade reading level books and samples is significantly smaller than those of even a few grades later. Most second graders will not likely be able to read words of more than a few syllables, which has a number of consequences. There are fewer common reading errors, and while pronunciation errors at the second grade level may highly differ from the correct pronunciation, children in later grades pronouncing longer words will make errors of a nature that is evident to any human listener yet difficult to distinguish to a computer. Also, the authors mentioned that they used a list of 52 words that were often mis-recognized by Sphinx-II yet pronounced correctly by the children. Naturally, as children get older, this list is also likely to grow and become unmanageable.

Also, smaller children are likely to read much slower than children in older grades, even those with poor reading ability. A high level of dialect-specific pronunciation is likely to enter the speech of older children and trip up the speech recognition system. Finally, although Emily's sentence based approach is helpful for smaller children, older children are often asked to read passages as sentences are fairly short, so the speed increase is not necessarily present.

Another interesting observation by the authors is that the "potential" reading level of the second graders was significantly lower than expected. They attributed this deficiency to the fact that the children's attention wandered due to the lack of a human face in front of them. However, this issue touches upon more important issues with using computers for educational purposes. Many listening comprehension skills are aided by human interaction, as many important cues are obtained from facial expressions and lip movement as well as gesture and body language. Obviously, reading does not include human interaction, but the "potential" reading level does.

For future research, the researchers might consider using some machine learning to train the system to achieve the correct confidence level required for intervention of the reader. It seems that much of the system consists of intensive manual fine-tuning, which does not allow the system to scale very easily and also creates problems for porting the system to other languages and reading levels.

The most interesting aspect of the system is the way in which the authors of the system handle repetitions and restarts of words as well as modifying the probabilities of possible next words to create a bias towards the correct reading. Future research might add some statistical language models to aid in assessing the word-to-word transition probability to reduce the amount of false alarms.