SRL Newsletter

Volume 1, Number 2 March 2001

ModelTalker: Latest Developments

A Personalized Speech Synthesizer

The ModelTalker speech synthesis system wants to bring the state of the art in speech synthesis to the AAC community. With the system, people can create their own personalized voices with the quality of recorded natural speech for common utterances and high quality synthetic speech in the same "voice" for new and unusual utterances.

ModelTalker Speech Inventories:

The ModelTalker System requires that a speaker record an inventory of speech. The recorded inventory is converted to a database used by the speech synthesis system. The database must have at least one example of each speech sound. The more examples of a sound there are, the more the speech synthesis system has to choose from, and thus the better the quality of the resulting synthesized speech. However, for reasons ranging from limited access to a recording system to voice fatigue, most people are unable to record huge inventories of speech. Thus the

continued on page 2

INSIDE THIS ISSUE

1 **ModelTalker: Latest Developments**

1 **Text-to-phonemes: Converting text for the ModelTalker Synthesizer**

3 **Other Research Under Way**

Text-to-phonemes

Converting text for the ModelTalker Synthesizer

When ModelTalker (and other speech systems) turn text into speech, it must first convert text into its corresponding phonemes (the basic sounds that make up words). At first thought, the conversion seems like a very tenable problem. After all, most of us convert text to sounds quite easily. Upon further reflection, however, the problem becomes unwieldy. To convert text to phonemes, we use rules. These rules start out simply enough. For instance, 'b' is a phoneme in the word "boy" and "tub". However, what about 'a'? The 'a' in "rat" is a different phoneme from the 'a' in "rate". Our rules become more complex, but they're still not untenable. But then there's the 'a' in "wad" versus the 'a' in "was". Exactly what rule should we write that correctly translates the 'a' in each of these cases? And what about the 'th' in "though" and "thine" versus the 'th' in "thought" and "thief"? To make matters even more complicated, where we place stress in a word can affect how the word is pronounced. For instance, the second 'o' in "PHOT-o-graph" is pronounced differently than the 'o' in "phot-O-graph-y", and so is the 'a'. Rules must be written to handle the more subtle sound differences that occur for example with the initial 'i' in "INdigo" versus "inDEED" or the second syllable in "SYStem" versus "sysTEMic". Finally, there are words like "read", which in one context ("I like to read.") are pronounced quite differently than in another context ("She read a book.").

Clearly writing rules to convert text into its proper phonemic translation is a complex and difficult task. In some cases it isn't possible to write a rule for certain words. These words are handled by creating

Continued on page 3

We need your input!

In order to assess our progress and make sure we're on the right track, we need input from clinicians and people with ALS. We plan to hold a meeting at A.I. duPont Institute Hospital for Children on Wednesday, April 25th, at which we will demonstrate the ModelTalker system and discuss the current status of the project. From clinicians and people with an interest in the project, we hope to get feedback on the system, on our goals and priorities, and insights on how we might better accomplish these priorities.

If you are interested in attending the meeting, please contact James Polikoff at (302) 651-6843, or email polikoff@asel.udel.edu. Attendance will be limited so please let us know as soon as possible if you would like to participate.

You can make a difference in regards to the direction our research takes. ***We hope to hear from you!***

Continued from page 1

content of the inventory must be selected carefully in order to maximize the coverage of certain speech sounds while minimizing the amount of recording required from a speaker.

Current Inventory

The inventory currently consists of a set of user-selected words and phrases, 250 high-frequency words and phrases, the most common AAC phrases (according to the University of Nebraska-Lincoln's AAC vocabulary web site, url:http://aac.unl.edu/vocabulary.html), function words (e.g. "the", "and") in many different contexts, and another set of phrases that cover 90% of the sound pairs that aren't covered by the rest of the inventory. These phrases consist of 3-syllable nonsense and real words, longer nonsense sentences (included as long sentences because it takes less time to record 3 long sentences than 6 shorter sentences), and a set of short sentences covering between-word sound pairs.

Preliminary Refinements:

Currently the inventory takes approximately 3-6 hours to record. However, we suspect that there's still a fair amount of redundancy that can safely be eliminated from the inventory without affecting the quality of the resulting synthetic speech. In order to reduce the inventory size, we have recorded 3 adults over the age of 65, and 3 adults between the age of 18-45. (We were unable to use one inventory from an over-65 adult.) At the most recent ISAAC 2000 conference, we reported research on the other 5 inventories in which we selectively removed the long, nonsense sentences from the inventory and then synthesized the sentences we removed. We then reincorporated the sentences that, when synthesized, were the least natural, and left the most natural sentences out (assuming that they didn't need to be in the inventory because they were the most redundant). We repeated this process a number of times until there was only minimal change in the quality of the resulting synthesized sentences and found that it is, in fact, possible to remove certain sentences with only a minimal effect on the quality of the resulting synthesized speech.

What's Next with the Inventory:

While we have looked at whether redundancy exists in the long nonsense sentences, we have yet to evaluate the redundancy in any of the other sets of phrases in the inventory. The only set that is inviolate is the set of user-selected words and phrases. So our next step will be to evaluate the redundancy in other phrase sets in the inventory. Our goal is to group phrases in the inventory in such a way that we can tell potential recorders that they must record group 1 for minimal quality, and that recording each additional group should give them an x% increase in quality with the resulting synthetic voice. Thus users can select how much they are comfortable recording and whether they want to make the effort to record more speech.❖

Text-to-phonemes: continued from page 1

an "exception" file, which holds words and their phonemic translations. Since the speech synthesizer searches the exception file for all entered text words, we want to keep the size of the exception file relatively small to cut down on search time. In other cases, adding one rule affects the translation of a set of words that were already being translated correctly. In this case we must determine whether adding the rule or leaving it out results in more words being translated correctly.

These are the issues we are dealing with in developing a text-to-phoneme translation system. Dr. Steven Hoskins developed our rule system for text-to-phoneme translation. Since the early summer, Jason Lilley has been working diligently to improve our system. He started by comparing word pronunciations produced by our system against the pronunciations of about 129,000 words listed in the Carnegie Mellon Pronouncing Dictionary produced at Carnegie Mellon University. It was found that our system differed from the CMU pronunciations for about 10,000 common words. Those 10,000 words have been sorted into about 130 lists, each list consisting of words that fall into a certain pattern of mispronunciation. For example, one list contained incorrectly pronounced words that begin with "ex-" (should it be pronounced "eks-" as in "extra", or "egz-" as in "exact"?) Another list contained words in which the letter "s" was mispronounced (e.g., the 's' in "kerosene" and the 's' in "enclose"). Theoretically, each pattern should reveal one (or a few) new rules that can be added to the system to fix the pronunciation of these words. Since June, about 120 rules have been added to our system and only 150 exceptions to the system. The number of mispronounced common words has been reduced from 10,000 to about 6,500. We are optimistic that we can eliminate the remaining 6,500 by the end of the spring.❖

Other Research under way…

STAR: Speech Training Assessment & Remediation

Also in development at the Speech Research Laboratory is the STAR system – a tool designed to help young children overcome articulation problems. The system, designed to work in conjunction with regular therapy, is in the form of a game to make the task of practicing articulation more enjoyable for children, and hopefully to keep them practicing for longer periods of time. The game uses an animated Star travelling through space who runs into space monsters that, with the child's help, must be taught to speak. Under the guise of teaching the monster, the child is prompted for words containing the phoneme with which she is struggling. The words become increasingly difficult. When the child correctly articulates the prompted word, she receives positive feedback (e.g. lights, small monster parts, etc.) until the child is able to correctly articulate the most difficult words, at which point it is assumed she has mastered the phoneme.

Feedback Evaluation

We have just completed a preliminary evaluation of the accuracy of the feedback children receive from the STAR system. Recognizing and assessing the many variations of a particular phoneme that children produce is an important and difficult part of the STAR system. To make sure our system was accomplishing this task with acceptable accuracy, we collected data from 8 children between the ages of 4 and 7 who were receiving speech therapy at the duPont Hospital for Children for a particular phoneme ('r'). We then had 50 undergraduate students in the Speech and Hearing Department at the University of Maryland rate the goodness of the phoneme.* Finally, we used our speech recognition system to rate the goodness of the utterances and compared our results to those of the students. Overall, we found a correlation of .75 between our ratings and those of the students. Further, for the most crucial subset of the data (cases where the child's 'r' sounded like 'w'), we observed a correlation of almost .9 between our ratings and those of the students. Although future evaluations are needed, these results suggest our system is quite adequate for giving feedback on articulatory

continued on page 4

**We are grateful for the assistance of Professor Nan Bernstein Ratner in obtaining these data.*

Other Research Under Way: continued from page 3

quality.

What's Next

Our next immediate goal is to evaluate the overall effectiveness of the system as a training tool as well as its appeal to children. To evaluate the effectiveness of the STAR system, STAR will be used in conjunction with therapy by children who have difficulty with two of the phoneme pairs STAR is currently capable of training (r-w and t-k). For each phoneme, sometimes the child will receive STAR training in addition to standard speech therapy, and sometimes the child will receive standard therapy alone. By monitoring each child's progress throughout the experiment, we will be able to assess each child's rate of learning either with or without additional practice using the STAR system. We predict that children will acquire speech skills more rapidly when they receive STAR practice in conjunction with therapy. We will also get feedback from the children, their parents, and their therapists on what aspects of the game they enjoyed and what areas could use improvement. We will use this information to direct future improvements to the system. ❖

SPEECH RESEARCH LABORATORY

The SRL Newsletter is published biannually by the Speech Research Laboratory (SRL), of the duPont Hospital for Children and the Center for Applied Science and Engineering of the University of Delaware

Speech Research Laboratory
duPont Hospital for Children
1600 Rockland Road
Wilmington, DE 19803.

Phone: 302-651-6843
Fax: 302-651-6895
Email: speech@asel.udel.edu
http://www.asel.udel.edu/speech/

ACKNOWLEDGEMENT

The Speech Research Laboratory is funded by NIDRR and by the Nemours Research Programs.

If you are interested in receiving this newsletter in audio format, braille format, or large-print, call us at: (302) 651-6843 or email yarringt@asel.udel.edu

Speech Research Laboratory
duPont Hospital for Children
1600 Rockland Rd
Wilmington, DE 19803