

10 Innovative Speech Applications

Deborah A. Dahl

Conversational Technologies

1820 Gravers Road
Plymouth Meeting, PA 19462 USA

+1 (610) 888-4532

dahl@conversational-technologies.com

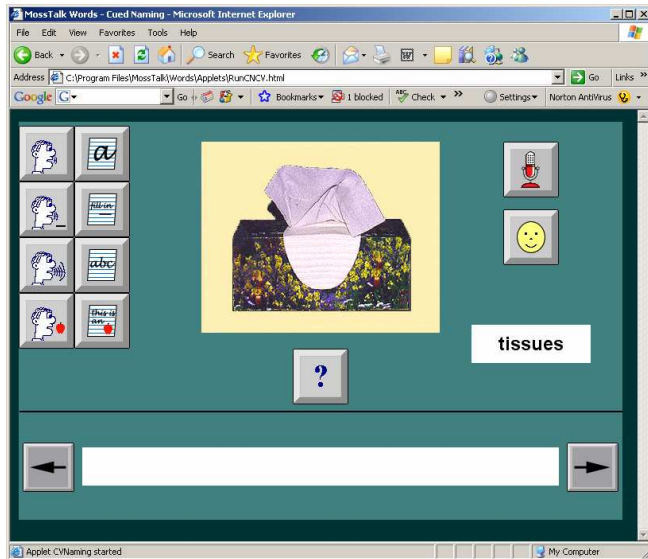
1. Introduction

Because of the success of speech applications involving telephone self-service and dictation, it's easy to get the impression that these are the only ways to apply speech technology. However, there are many applications of speech currently being developed by creative developers that go beyond these better-known applications. Looking at these applications provides a much richer picture of the many ways that speech technology has the potential to improve people's lives. The applications that will be discussed include applications of both speech recognition and text-to-speech, as well as other analyses of the speech signal. They include projects at all stages of development, from advanced research projects to commercial products. One characteristic that they all share is that in general, they don't rely on highly advanced speech technology itself; rather, they all illustrate what can be done with today's technology applied in innovative ways. The goal of this paper is to make readers more aware of the potential applications of speech technology, and perhaps spark some ideas for new innovations.

2. MossTalk Words

Over one million Americans have difficulty understanding or producing language because of an injury to the parts of the brain that control language, most commonly caused by a stroke. This condition is called *aphasia*. Depending on the severity of the condition, aphasia can be extremely socially isolating, and often prevents people from being able to work. Although in many cases insurance pays for a period of speech therapy, typically insurance reimbursement for speech therapy is limited, even when the patient could potentially benefit from additional speech therapy.

MossTalk Words, developed by Moss Rehab hospital in Philadelphia, provides automated assistance with one type of speech therapy by helping people with aphasia practice finding and speaking words. The user is presented with a picture of a common object, and tries to speak the word corresponding to the object. Originally, a clinician working with the user provided feedback on whether the user's utterance was correct. However, the system is currently being modified so that the feedback can be provided by the system itself using speech recognition. If the user says the wrong word, or mispronounces the word, nothing happens. However, if the user speaks the word correctly, the system says "That's right", and says the word again. The picture shows the screen the user sees after correctly speaking the word "tissues".



Some advantages of using speech technology for speech therapy include lower cost, 24/7 availability, consistent responses, and automatic recording of the users' performance.

More information can be obtained from www.mosstalk.com

3. English X-Change

There are estimated to be over 225 million people who would like to learn English in China, but there are only 100,000 English teachers, most of them not native speakers. These teachers are also concentrated in urban areas; consequently it is even more difficult to learn English in rural areas.

English X-Change has developed software that provides a simulation-based, interactive computer program to help Chinese students learn English. Many of the lessons require students to speak English. The English Exchange software program uses speech recognition to evaluate students' pronunciation and provide feedback to the students. The degree to which a student's pronunciation of a word or phrase approaches correct native English pronunciation can be adjusted.

The English instruction provided by English X-Change has been shown to be extremely effective. In one study, students who studied using the English Exchange software program produced substantially and significantly higher test scores than did those who experienced traditional classroom instruction with trained native English speakers.

More information can be obtained from www.englishexchange.com.

4. Compliance for Life

People don't always take their medications as directed. According to Jane Brody of the New York Times, "The misuse or non-use of prescribed medications is estimated to add nearly \$200

billion a year to the cost of medical care.” (May 9, 2006). Non-compliance rates can be very high. For example, hypertension non-compliance is estimated to be around 40%. The most common reason that people give for not taking their medication is that they simply forgot.

Compliance for Life addresses this problem by providing a phone and web-based automated notification system to create, edit and cancel medication reminders. Reminders can be created by the patient, or by a family member. The phone interface allows users to use speech recognition to manage reminders when the web isn't available.

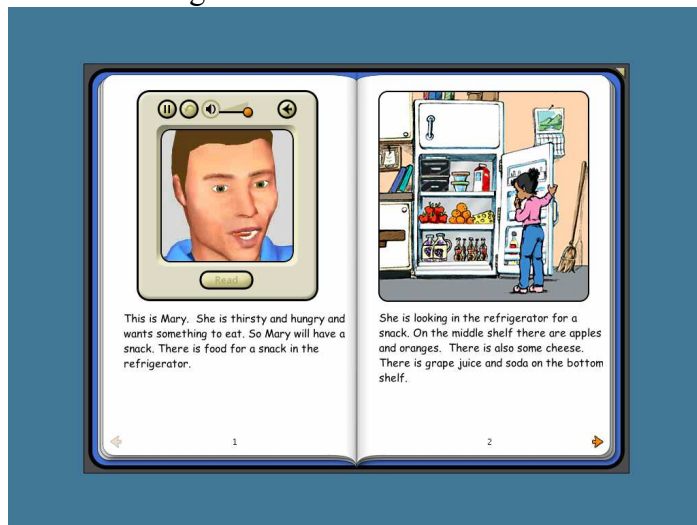
More information can be obtained from www.iReminder.com.

5. Animated Speech

Millions of preschool and elementary school children have language and speech disabilities. There is a shortage of skilled teachers and professionals to give them the one on one attention that they need.

Animated Speech applies animated agents to produce accurate visible speech and facilitate face-to-face oral communication. The Timo Stories product helps students with story comprehension, vocabulary, syntax, and story re-telling. The Timo Vocabulary product teaches vocabulary to children with language challenges. Instruction is always available to the child, 24 hours a day, 365 days a year. The automated system also has the advantage of being extremely patient, and doesn't become angry, tired, or bored.

The figure shows a page from Timo Stories, along with Timo, the animated tutor who guides students through the exercises.



More information can be obtained from www.animatedspeech.com

6. Diagnosing Depression

Depression is traditionally diagnosed and monitored by self-report—asking the patients how they feel. Some depressed individuals are reluctant to admit that their medication isn't working, and so may give overly optimistic reports of how they feel. However, speech from depressed patients can provide an objective measure of the severity of depression because depressed people's speech also differs from other people's speech in pitch, loudness, speech rate and articulation. For example, people who are depressed show decreases in pitch variability, more pauses, and a slower speaking rate. Healthcare Technology Systems has developed a system that automatically measures some of the acoustic correlates of depression severity in patients' speech. More information can be found at www.healthtechsys.com/ivr/ivrmain.html

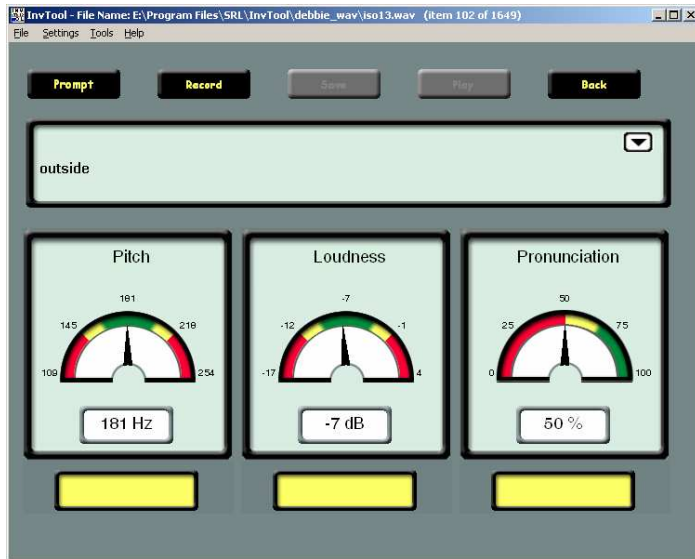
7. Guided Speech

Automatic speech recognition is widely used for telephone self-service, but it's not always accurate. Speech recognition failures can easily lead to extremely frustrating experiences for callers. Human speech recognition is accurate, but humans are expensive and get bored with handling routine calls. Spoken Communications Guided Speech IVR uses humans to back up speech recognition. In this system a human guide in the background assists self service application by listening to and handling problematic utterances without actually getting on line with the customer. Using this approach, agents are able to handle 4 calls silently and simultaneously.

More information can be obtained from www.spoken.com.

8. Model Talker

People who have limited ability to speak, or have lost their ability to speak, for example, as a result of ALS, often use TTS to speak their typed utterances. Concatenative TTS is much more intelligible than format-based TTS systems such as DECTalk. However, the number of available concatenative TTS voices is limited, and there may not be an existing voice that the user likes. Model Talker, developed by the Nemours Speech Research Laboratory located at the Alfred I. duPont Hospital for Children, lets users generate a TTS voice from their own recordings. If the user has lost the ability to speak altogether, he or she can select someone else to record a voice. The user uses a software tool to record a carefully selected inventory of sentences, phrases, and isolated words designed to cover almost all of the different combinations of speech sounds found in naturally occurring English. After recording, users then create a TTS voice by uploading their recordings to a voice generation site. In addition to people who have lost the ability to speak, this application is also of interest to blind users who want to record a voice for their screen readers. The picture illustrates the recording tool as it prompts the user to say the word "outside", giving the user feedback on pitch, loudness, and pronunciation.



More information can be obtained from www.modeltalker.com.

9. ASL Speech Recognition

American Sign Language is the fourth most-used language in the United States. Currently, human ASL translators are frequently necessary to facilitate communication between deaf and hearing presenters and their audiences; however, good ASL translators are in high demand and are not always available, which makes communication among deaf and hearing people quite difficult.

The goal of the ASR Project at DePaul University is to combine speech recognition and understanding with automatic ASL generation to translate from spoken English to American Sign Language. This application combines speech recognition, natural language understanding, and ASL synthesis.



More information can be obtained from asl.cs.depaul.edu.

10. VoiceBox In-Car Navigation

Current in-car navigation systems require multiple button presses to set destinations. In one test of Neverlost the average time to set a destination for 25 first time users of was 4 minutes and 31

seconds, and required 315 button pushes. In this test, five testers dropped out and said that they could not do it.

For setting destinations in an in-car navigation system, speech is much faster and less confusing. For the same task of setting a destination, the VoiceBox time average was 18 seconds for new users.

Here's an example dialog between a user and the VoiceBox navigation system.

User: Show McCarran International Airport.

System: Showing McCarran International Airport.

User: Set this as my starting point.

System: Where in Las Vegas would you like to go?

User: Set a course for Wynn Las Vegas.

System: (displays map) Directions to Wynn Las Vegas.

User: Show destination.

System: Destination display.

User: Show nearby Italian restaurants.

System: Restaurants, Italian points of interest.

User: Select number 4.

System: Ristorante Italiano.

User: Show McCarran International airport.

System: Showing McCarran International Airport.

User: Cancel route.

System: Route canceled.

User: Show nearby Starbucks.

System: Here's the nearest Starbucks.

User: Could you call them?

System: Hold on a moment while I set up a hands-free call to Starbucks.

More information is available from www.voicebox.com.

11. Rex the Talking Pill Box

Some patients can't read or understand the instructions on their prescription bottles, due to such factors as illiteracy, low vision or cognitive limitations. Currently, the only way that these patients have to understand their medication instructions is to either find someone to help them with the instructions or simply try to remember the instructions. Either of these alternatives could lead to the patient either taking the wrong medication or not taking their medication according to instructions. Rex the Talking Pill Box provides an inexpensive voice solution to helping patients who can't read the instructions on their prescription bottles.

When a prescription is filled using the Talking Pill Box, the pharmacist programs the bottle with the medication instructions, which are rendered with TTS when a button is pushed. Alternatively, users with the home version of the system can record their own messages.

More information is available from www.rxtalks.com.

12. Summary

This paper has reviewed ten innovative applications of speech technology, including applications of speech recognition, text to speech and other analyses of the speech signal. The applications range from university research projects through deployed commercial products

Several themes run through these applications:

- Speech technology can improve a graphical user interface that is clumsy and difficult to use (VoiceBox navigation).
- Speech technology can compensate for expensive or unavailable human expertise (MossTalk, Timo Stories, English X-Change, ASL translator, depression diagnosis, and Guided Speech).
- Speech technology can help users who need assistance
 - understanding or producing spoken language (ASL translator, Model Talk)
 - remembering (Compliance for Life)
 - understanding or producing written language (Rex, Timo Stories)