CONTACT: John Toon/Ginger Pinholster
(404) 894-3444

TALKING TO MACHINES: PATTERNS IN KEY COMPONENT OF SPEECH MAY HELP RECOGNITION OF "STRESSED" SPEECH

For Immediate Release

June 11, 1990

Photography Available

Researchers studying a key component of human speech patterns have found important consistencies that could help automated speech recognition devices understand different persons across a wide range of speaking styles. The work could improve automated speech recognition, processing and synthesis techniques by expanding knowledge of how emotions affect speaking patterns.

Automated speech processing offers tremendous potential for increasing man-machine interaction, but variations in speech patterns due to individual differences, environmental factors -- and the effect of emotions like anger and fear -- have severely limited its use.

At the Georgia Institute of Technology, researchers are studying the component of speech produced by the periodic opening and closing of the vocal cords at the upper part of the larynx. Called the glottal waveform, this opening and closing provides excitation to generate sound, which is then filtered by the vocal tract and shaped by the mouth and lips to form the words humans use to communicate.

Researcher Kathleen Cummings and Electrical Engineering Professor Dr. Mark Clements studied how the glottal patterns changed under "stressful" conditions. Stressed speech can occur, for instance, as persons become angry, attempt to make themselves heard over noisy backgrounds or are distracted by difficult tasks.

"We have found some consistencies across different speakers and different utterances," said Clements. "There seems to be a predictable pattern of modification -- certain things that happen to the glottis during these times."

An understanding of how speech patterns are altered could be the basis for a process to automatically convert emotionally-charged speech to a standard form more easily recognized by machines, Clements suggested. The work could also lead to more natural synthetic speech, and potentially to new devices which would automatically clarify speech for better understanding of critical communications.

- OVER -

Speech recognition devices must be "taught" what each combination of sounds means. Because each person produces slightly different sound patterns when speaking the same words, the most reliable devices are trained to a particular individual, leaving them "speaker dependent."

Another critical problem is that changes in a person's emotional state or factors in the environment affect the way a person speaks the same words. These variations alter the sound patterns, and can confuse speech recognition devices.

"If you are in a noisy environment, you tend to talk differently," explained Clements. "It changes the recognition a lot not only because of the background noise, but also because of the way the person alters his speech in the presence of noise."

Automated speech recognition could be particularly valuable to military pilots and others who must handle many tasks simultaneously. Pilots in the heat of battle could one day use speech recognizers to convey commands -- but only if recognizers can adapt to changing emotions and confusing background noise.

"The times when you would want it to have the highest accuracy are the worst times for speech processing," Clements noted. "It's hard to predict how a pilot's speech would change when he suddenly sees a mountain in front of him."

Cummings and Clements believe these emotionally-dependent changes come primarily from the larynx. By studying the glottal sound waves from 11 different types of speech, they have characterized the changes and hope to predict how a person's sound waves would vary under different conditions.

"It opens the possibility of speaker independent recognition," said Cummings. "Any recognition system, to be practical, is going to have to understand and take into account these sorts of effects."

The researchers used computerized signal processing techniques to isolate the glottal component of recorded words spoken by a variety of male speakers.

The researchers ultimately hope to separate three components of speech: the word meaning, the identity of the speaker, and the emotional state of the speaker. Filtering out the individual components could give speech recognizers tremendous capabilities.

Despite considerable attention, Clements believes that speech recognition technology remains in its infancy, with available devices still not practical for widespread applications. To be truly practical, scientists must make the devices speaker independent, able to recognize a wide range of speech types -- and able to recognize continuous speech uttered without separation between words.

Improvements will come incrementally as more is learned about speech recognition -- and more computer power becomes available, he said.

The work has been sponsored by the U. S. Army's Human Engineering Laboratory, the National Science Foundation and Georgia Tech research funds. Information on the work has been published in the proceedings of 1989 IEEE Southeastcon and the proceedings of the 1990 IEEE International Conference on Acoustics, Speech and Signal Processing.