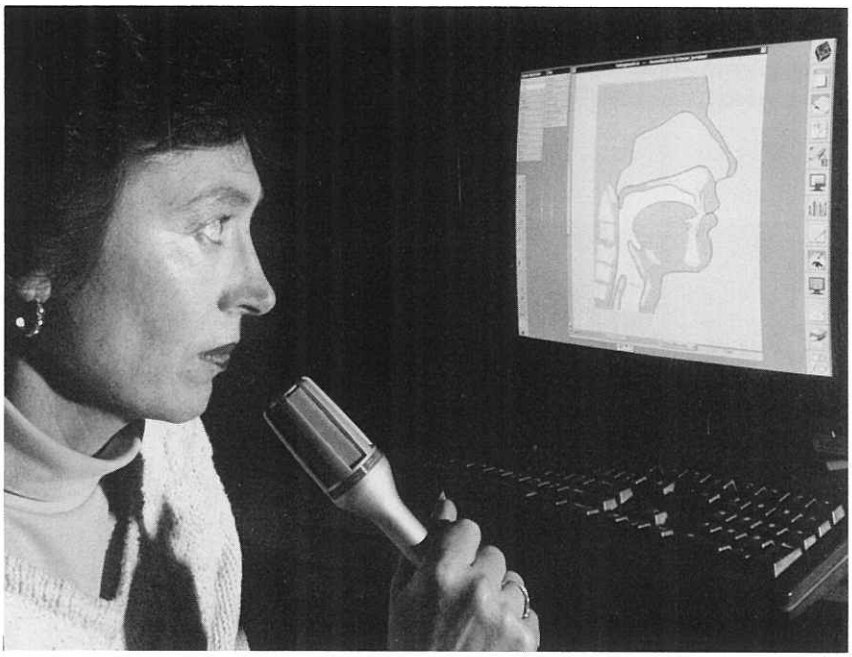
Animated Display of Inferred Tongue, Lip, and Jaw Movements During Speech

Inventors: George Papcun, Timothy Thomas, and Judith Hochberg, Computing and Communications Division



RN92-048 002

As the speaker talks into the microphone, the computer's artificial neural network generates moving graphic images of her lips, tongue, and jaw movements. The computer system generates the movements by relating the speaker's sounds to the movements in the parts of the vocal tract that produce those sounds.

Young children sometimes have trouble learning to speak and need speech therapy. Among the most difficult sounds for therapists to teach are the /g/ and /k/ consonants because they are formed by the tongue in the back of the throat and the child cannot see the tongue's movements.

Researchers at Los Alamos National Laboratory have developed a technology that can serve as the basis for a new system. With this system, speech therapists could "show" children and other patients how to speak correctly. As the patient talks into a microphone attached to a specially equipped personal computer, the computer's display screen would show a simulated x-ray motion picture of the movements of the tongue, lips, and jaws—the structures of the vocal tract. The computer would also display correct vocal tract movements for the patient to imitate. In addition, the system could be used to correct accents and to teach the deaf to speak.

Because of these potential applications, our technology won a 1992 R&D 100 Award, presented annually by *Research and Development Magazine* to the one hundred most significant technical innovations of the year. Los Alamos has a patent pending on the device. In recognition of its potential for providing assistance to disabled persons, our technology was displayed in February 1992 at the Smithsonian Institution in Washington, D.C., as one of the thirty winners of the Johns Hopkins University National Search for Computing to Aid Persons with Disabilities.

The Invention—Characteristics and Advantages

Our speech analysis technology is based on an artificial neural network that infers the movements of a speaker's vocal tract from the acoustical input it receives from the speaker. We "trained" the neural network using real speech data—speech acoustics and the corresponding movements of the structures of the vocal tract.

To collect these data, we used the University of Wisconsin's x-ray microbeam facility to obtain recordings of a group of speakers. After attaching tiny gold pellets to a speaker's lips, tongue, and jaws, we directed an x-ray microbeam at the pellets. The pellets absorbed energy from the x-ray, and a computerized sensor system recorded the movements. Using this technique, we measured the movements of the lips, tongue, and jaws in terms of the position, velocity, and acceleration of the pellets.

This process, repeated with many speakers, resulted in a set of records consisting of speech acoustics and the corresponding movements that produced the speech. The records were then used to train the artificial neural network to map, or relate, the acoustics to the corresponding movements.

To use a speech therapy system based on our technology, a person would speak in a normal tone into a microphone attached to a personal computer equipped with high-resolution graphics. The neural network would analyze the acoustic properties of the speech input and use the learned map to infer the speech movements that produced the input. The computer would display a simulated x-ray photograph of the speaker's vocal tract and its movements.

Applications

A system based on our technology could be used to help treat speech disorders, to teach proper accents in foreign languages, and to modify regional dialects in English. In each of these applications, the speaker would attempt to reproduce the same articulatory motions as those of a model speaker.

We anticipate that our technology can also serve as a starting point for other applications. Existing speech

For more information about the method for Animated Display of Inferred Tongue, Lip, and Jaw Movements During Speech, please contact Kay Adams, Industrial Partnership Center, Los Alamos National Laboratory, P.O. Box 1663, Mail Stop K763, Los Alamos, NM 87545, Telephone (505) 665-9090, Fax (505) 665-3164.

LALP 92-64
September 1992

recognition systems can identify only isolated words and phrases with reasonable accuracy; when these systems attempt to recognize continuous speech, they are much less accurate. Because computerized speech recognition would open up many applications, the federal Office of Science and Technology Policy identified computerized speech recognition as one of twenty "grand challenges" for modern computational science.

Our technology represents a crucial step toward meeting this challenge. Conventional approaches to speaker-independent speech recognition consider speech as a sequence of discrete units. By contrast, our technology represents speech as overlapping constellations of articulatory gestures, or concrete physical movements. Because the technology relates speech acoustics to the physical movements that produce those sounds, it can lay the groundwork for continuous speech recognition systems.

When computers can recognize fluent speech accurately, computer users will no longer be restricted to keyboard entry of data or commands; typing will be reduced or eliminated as part of the office routine; and the public's access to information stored by computers and to operations governed by computers will be greatly increased.

George Papcun is the speech project leader in Los Alamos National Laboratory's Computing and Communications Division; he has been a staff member at the Laboratory since 1982. After earning a Ph.D. in linguistics at the University of California at Los Angeles, he pursued research in computational linguistics and in how humans perceive speech and language. Papcun has served as an expert witness in legal cases involving voice identification and the intelligibility and fabrication of recordings. He has consulted with law enforcement agencies on projects involving voice identification. He has also taught engineering courses on computerized speech synthesis and recognition.

Timothy Thomas came to the Laboratory as a research affiliate in 1986 from a fifteen-year career in college and community college teaching. In 1989 he became a staff member in the Computer Research Group of the Communications and Computing Division. Currently, he is pursuing research interests in pattern classification and anomaly detection. His M.S. in psychology and Ph.D. in physiological psychology are from Tulane University.

Judith Hochberg has been a staff member in the Computer Research Group of the Computing and Communications Division since 1991; she came to the Laboratory in 1989 as a director-funded postdoctoral staff member. Her research for her M.A. and Ph.D. in linguistics at Stanford University focused on the way children master the sound system of language. Her work for the speech recognition project spurred an interest in the field of computational linguistics in general and anomaly detection in large data bases in particular.

RB 92-055 004

Inventors of the method for Animated Display of Inferred Tongue, Lip, and Jaw Movements During Speech are (from left) Timothy Thomas, Judith Hochberg, and George Papcun.

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the US Department of Energy under contract W-7405-ENG-36.