

# Seeing Speech, Perfecting Parlance

BY JAMES SCHULTZ

“Aaaah”

“Ooh”

“Eeee”

“Uhhh”

ee

ue

lh

oo

ur

ch

uh

aw

ae

ah

Stephen Zahorian settles his lanky frame into an office chair, untangling lightweight headphones that snake into a nearby computer, slipping the phones over his head and adjusting the small, attached microphone. He leans back, speaking slowly and distinctly. “Ooh,” Zahorian intones. “Aaaah. Eeee. Uhhh.”

Swirling at eye level on the screen in front of Zahorian are ellipses of green, turquoise, yellow, purple, navy blue and red. When Zahorian manages to produce the correct sound, a thick dot of color blossoms sun-like inside each vowel-specific orbit. It’s no accident that the exercise is designed like a video game: not only fun to play, but effective in assessing and tracking a user’s language-matching progress.

Zahorian, professor and chairman of Old Dominion’s Department of Electrical and Computer Engineering, calls his system a “visual speech training aid,” a unique concatenation of advanced software and multimedia-enabled hardware that represents more than a decade’s worth of effort in understanding and replicating the intricacies of spoken human language. Zahorian’s goals are threefold: to automatically analyze sounds and words as they are spoken, to monitor correct pronunciation and to provide an ongoing means for speakers to improve through self-correction.

Although geared toward those who are hearing impaired, the visual-speech system also has long-term potential as a translation device. It may also one day be used to teach English as a second language or even assist hearing individuals who wish to practice or perfect locution.

## Demand And Motivation

Zahorian estimates there are currently some 250,000 hearing-impaired individuals in the United States between the ages of 5 and 21. While the incidence of deafness in school-age children has remained fairly constant, roughly three in every 4,000, the rates of mild-to-severe hearing loss are almost seven times higher, about 20 in 4,000. In addition, within the hearing-impaired community itself, a growing number of individuals is receiving cochlear implants and so require some form of speech training. The demand may be even greater in the hearing community; normal-hearing children with speech-articulation disorders could stand to benefit from the visual-speech system.

While in theory acquiring or improving speech can be and is done with the help of an experienced therapist, in practice speech therapy is expensive, time-consuming, and the supply of trained therapists-limited. Because Zahorian’s proposed system is installable on most personal computers, the program would be constantly available for practice. As with any extended effort, mastery isn’t guaranteed; progress

would depend on frequency of use.

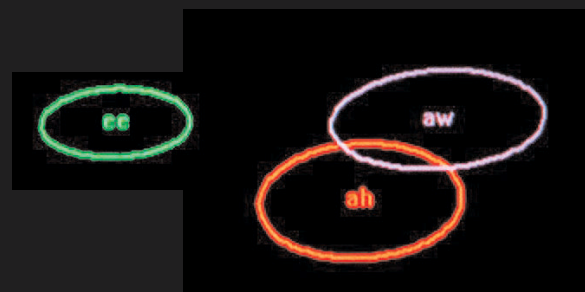
“If you’re hearing impaired, learning to speak orally is extremely difficult,” Zahorian says. “While sign language works and works quite well, this system provides an extra tool to those who are hearing impaired and who may desire to acquire oral speech. You can talk to a computer as many times as you want without fear of embarrassment. I’m convinced that anyone using this system would be very motivated to learn to speak.”

Zahorian’s interest in speech recognition was kindled by a longstanding fascination with signals processing and the difficulty in devising effective pattern-recognition programs to identify complex signals. As an engineer, he remains intrigued by the apparent simplicity of speech that masks an innate and daunting intricacy: Even though certain frequencies and certain components figure prominently in human language patterns, isolating and exactly matching those patterns is a challenging and time-consuming process.

In the current iteration of the visual-speech program, Zahorian has compiled a database of spoken language from approximately 300 male and female adults and children. Each speaker was asked to pronounce vowels, vowel-consonant combinations and short consonant-vowel-consonant words. Listeners monitored each recording session to ensure that utterances were intelligible. Software captured the resultant vocal signals, analyzing and caching the signals’ unique digital formulations. Zahorian, with substantial assistance from Old Dominion graduate students Neiyer Correal, Stefan Augerg, Xihong Wang, Benjamin Dai and Matt Zimmer, subsequently developed a computer-based neural network — modeled after the ways in which the human brain processes and makes sense of new information — to further process and display the sounds in a visual format.

“None of this would be possible without the computer revolution,” Zahorian says. “That’s what’s powering the possibilities in speech recognition. Engineers like me keep trying to integrate what we’ve learned in signals processing and speech science. We’re making steady progress. But the average 3-year-old kid still understands speech better than the best software program.”

The system represents more than a decade's worth of effort in understanding and replicating spoken human language



# Improving System Capabilities

Although Zahorian says that he and other of his speech-recognition colleagues have made significant progress in reproducing human speech, natural language for computers remains difficult. In part it is because speech itself is composed of many disparate parts, including vocalization, that work together synergistically. Making sense of speech is therefore daunting in ways that other, less challenging computations are not. That human beings are so skilled at language acquisition and use has much to do with the brain's massive interdependence and the sheer computing ability of a trillion-plus neurons and neuronal interconnections.

Depending on progress made in the next phase of the visual-speech research — in 1999 the National Science Foundation provided \$90,000 each year for three years to Zahorian to continue the system's development — a version of the visual-speech program could become commercially available within five years. A more rudimentary form of the visual-speech system was tested in the Chesapeake, Virginia school system, with promising results. Currently, Zahorian estimates his speech program is "about 85 percent accurate," a figure he

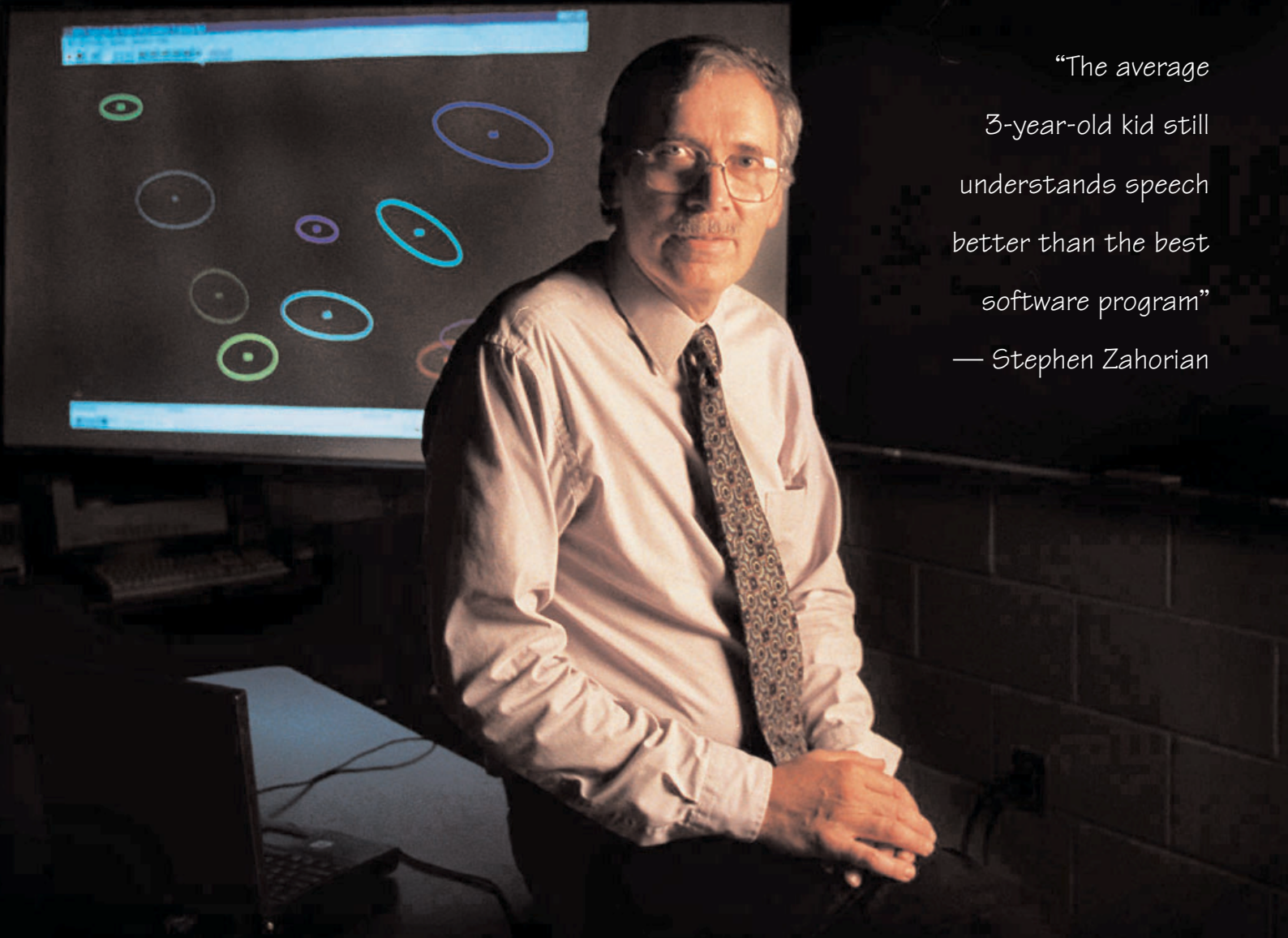
intends to substantially increase over the next several years.

"One reason I don't have it yet out in the public schools is that we're trying to improve the accuracy to where people aren't discouraged by the computer's mistakes," he says. "I'd also like to incorporate some computer games, where you control the action with the vowel sounds you make.

"I want to make the system speaker-independent, able to discern the correct phonetic response despite the frequency range. We want to guarantee the system will work whether a person's vocal timbre is high, low or somewhere in the middle. The long-range goal is to produce something that is really useful to anyone who wants to learn oral speech."

In the immediate future, Zahorian will be enlarging his speech database. Spoken language samples will be collected from another 300 speakers with normal hearing, 100 each of adult men, adult women and children under the age of 10. Within each group, a broad collection of regional dialects from throughout the United States will be included. Hearing-impaired individuals will also participate.

In addition, Zahorian envisions further improvements to computer software and hardware. By the end of calendar year 2000, experiments with foreign-language training are planned. Testing of the system's improved capabilities with hearing-impaired speakers is slated to occur by the end of 2001.



*"The average  
3-year-old kid still  
understands speech  
better than the best  
software program"*  
— Stephen Zahorian