

July 23, 2007

Bare-Bones Program Learns English and Japanese Vowels

Computer model learns vowel sounds infant-style: on the fly

A new computer model has learned to recognize vowel categories from multiple English and Japanese speakers without "knowing" the number of vowels it is looking for or having a complete list of sounds to analyze, according to a new report. Instead, it gradually lumps vowels into distinct groups by considering them one at a time, reminiscent of how an infant might attend to sounds.

The designers of the model say it is an early step toward improved voice recognition software and a better understanding of how the infant mind comes to recognize that the voices it detects are speaking one language and not another.

"We see this work as representing a movement towards thinking about language learning as an experience-dependent process," says James McClelland, professor of psychology at Stanford University and co-author of the report appearing online in *Proceedings of the National Academy of Sciences USA*.

Psychologist Janet Werker of the University of British Columbia in Vancouver recorded mothers in a laboratory speaking nonsensical sounds in English or Japanese to infants [[click here for sound samples](#)]. Both languages have five vowels that, roughly speaking, come in a long and short form, such as the English "bait" and "bet," which differ in frequency, whereas the Japanese vowels differ in the duration of the sound.

Trying to distinguish the "i" and "e" vowel forms for each language, McClelland, Werker and their colleagues converted each recorded vowel sound into three numbers that represented the duration of the sound and its two dominant frequencies. Then they fed these values into their model one vowel at a time.

The program placed each value on a continuum of many durations or frequencies that might possibly define a spoken vowel. The values reinforced particular durations or frequencies, gradually building up a three-dimensional space for each vowel form.

After such training, the program correctly categorized up to 93 percent of vowels in English and 92 percent in Japanese, the group reports.

McClelland says prior language learning models were less realistic, because they repeatedly scanned a large set of sound data instead of one sound at a time.

Incorporating similar procedures might allow speech recognition software to adapt to different speakers of the same language and thus boost its accuracy, he adds.

He says the new model is hard to compare with infant learning because researchers don't know what sounds infants hear. "But," he adds, "it's pretty successful at what it does and it uses a set of principles we think are on the right track."

**Does
Tea
Really
Have
More
Caffeine
Than
Coffee?**



ADVERTISEMENT

c|net CNET TV

Xbox 360 VS Sony PS3

Watch the tech showdown

WATCH VIDEO

PRIZEFIGHT

© 1996-2007 Scientific American, Inc. All rights reserved.
Reproduction in whole or in part without permission is prohibited.