Doing what the brain does -- how computers learn to listen

August 14 2009

(PhysOrg.com) -- We see, hear and feel, and make sense of countless diverse, quickly changing stimuli in our environment seemingly without effort. However, doing what our brains do with ease is often an impossible task for computers. Researchers at the Leipzig Max Planck Institute for Human Cognitive and Brain Sciences and the Wellcome Trust Centre for Neuroimaging in London have now developed a mathematical model which could significantly improve the automatic recognition and processing of spoken language. In the future, this kind of algorithms which imitate brain mechanisms could help machines to perceive the world around them.

Many people will have personal experience of how difficult it is for computers to deal with spoken language. For example, people who "communicate" with automated telephone systems now commonly used by many organisations need a great deal of patience. If you speak just a little too quickly or slowly, if your pronunciation isn't clear, or if there is [background noise](), the system often fails to work properly. The reason for this is that until now the computer programs that have been used rely on processes that are particularly sensitive to perturbations. When computers process language, they primarily attempt to recognise characteristic features in the frequencies of the voice in order to recognise words.

"It is likely that the brain uses a different process", says Stefan Kiebel from the Leipzig Max Planck Institute for Human Cognitive and Brain Sciences. The researcher presumes that the analysis of temporal

sequences plays an important role in this. "Many perceptual stimuli in our environment could be described as temporal sequences." Music and spoken language, for example, are comprised of sequences of different length which are hierarchically ordered. According to the scientist's hypothesis, the brain classifies the various signals from the smallest, fast-changing components (e.g., single sound units like "e" or "u") up to big, slow-changing elements (e.g., the topic).

The significance of the information at various temporal levels is probably much greater than previously thought for the processing of perceptual stimuli. "The brain permanently searches for temporal structure in the environment in order to deduce what will happen next", the scientist explains. In this way, the brain can, for example, often predict the next sound units based on the slow-changing information. Thus, if the topic of conversation is the hot summer, "su…" will more likely be the beginning of the word "sun" than the word "supper".

To test this hypothesis, the researchers constructed a [mathematical model](link) which was designed to imitate, in a highly simplified manner, the neuronal processes which occur during the comprehension of speech. Neuronal processes were described by algorithms which processed speech at several temporal levels. The model succeeded in processing speech; it recognised individual speech sounds and syllables. In contrast to other artificial speech recognition devices, it was able to process sped-up speech sequences. Furthermore it had the brain's ability to "predict" the next speech sound. If a prediction turned out to be wrong because the researchers made an unfamiliar syllable out of the familiar sounds, the model was able to detect the error.

The "language" with which the model was tested was simplified - it consisted of the four vowels a, e, i and o, which were combined to make "syllables" consisting of four sounds. "In the first instance we wanted to check whether our general assumption was right", Kiebel explains. With

more time and effort, consonants, which are more difficult to differentiate from each other, could be included, and further hierarchical levels for words and sentences could be incorporated alongside individual sounds and syllables. Thus, the model could, in principle, be applied to natural language.

"The crucial point, from a neuroscientific perspective, is that the reactions of the model were similar to what would be observed in the human brain", Stefan Kiebel says. This indicates that the researchers' model could represent the processes in the brain. At the same time, the model provides new approaches for practical applications in the field of artificial speech recognition.

More information: Stefan J. Kiebel, Katharina von Kriegstein, Jean Daunizeau, Karl J. Friston; Recognizing sequences of sequences; *PLoS Computational Biology*, August 14th, 2009.

Source: Max-Planck-Gesellschaft (news : web)