Computational model decodes speech by predicting it

June 26 2020

The brain analyzes spoken language by recognizing syllables. Scientists from the University of Geneva (UNIGE) and the Evolving Language National Centre for Competence in Research (NCCR) have designed a computational model that reproduces the complex mechanism employed by the central nervous system to perform this operation. The model, which brings together two independent theoretical frameworks, uses the equivalent of neuronal oscillations produced by brain activity to process the continuous sound flow of connected speech.

The model functions according to a theory known as predictive coding, whereby the brain optimizes perception by constantly trying to predict the sensory signals based on candidate hypotheses (syllables in this model). The resulting model, described in the journal *Nature Communications*, has helped the live recognition of thousands of syllables contained in hundreds of sentences spoken in natural language. This has validated the idea that neuronal oscillations can be used to coordinate the flow of syllables we hear with the predictions made by our brain.

"Brain activity produces neuronal oscillations that can be measured using electroencephalography," says Anne-Lise Giraud, professor in the Department of Basic Neurosciences in UNIGE's Faculty of Medicine and co-director of the Evolving Language NCCR. These are electromagnetic waves that result from the coherent electrical activity of entire networks of neurons. There are several types, defined according to their frequency. They are called alpha, beta, theta, delta or gamma waves. Taken individually or superimposed, these rhythms are linked to different cognitive functions, such as perception, memory, attention, alertness, etc.

However, neuroscientists do not yet know whether they actively contribute to these functions and how. In an earlier study published in 2015, Professor Giraud's team showed that the theta waves (low frequency) and gamma waves (high frequency) coordinate to sequence the sound flow in syllables and to analyze their content so they can be recognized.

The Geneva-based scientists developed a spiking neural network computer model based on these physiological rhythms, whose performance in sequencing live (on-line) syllables was better than that of traditional automatic speech recognition systems.

The rhythm of the syllables

In their first model, the theta waves (between 4 and 8 Hertz) made it possible to follow the rhythm of the syllables as they were perceived by the system. Gamma waves (around 30 Hertz) were used to segment the auditory signal into smaller slices and encode them. This produces a "phonemic" profile linked to each sound sequence, which could be compared, a posteriori, to a library of known syllables. One of the advantages of this type of model is that it spontaneously adapts to the speed of speech, which can vary from one individual to another.

Predictive coding

In this new article, to stay closer to the biological reality, Professor Giraud and her team developed a new model where they incorporate elements from another theoretical framework, independent of the neuronal oscillations: "predictive coding."

"This theory holds that the brain functions so optimally because it is constantly trying to anticipate and explain what is happening in the

environment by using learned models of how outside events generate sensory signals. In the case of spoken language, it attempts to find the most likely causes of the sounds perceived by the ear as speech unfolds, on the basis of a set of mental representations that have been learned and that are being permanently updated," says Dr. Itsaso Olasagasti, computational neuroscientist in Giraud's team, who supervised the new model implementation.

"We developed a computer model that simulates this predictive coding," explains Sevada Hovsepyan, a researcher in the Department of Basic Neurosciences and the article's first author. "And we implemented it by incorporating oscillatory mechanisms."

Tested on 2,888 syllables

The sound entering the system is first modulated by a theta (slow) wave that resembles what neuron populations produce. It makes it possible to signal the contours of the syllables. Trains of (fast) gamma waves then help encode the syllable as and when it is perceived. During the process, the system suggests possible syllables and corrects the choice if necessary. After going back and forth between the two levels several times, it discovers the right syllable. The system is subsequently reset to zero at the end of each perceived syllable.

The model has been successfully tested using 2,888 different syllables contained in 220 sentences, spoken in natural language in English. "On the one hand, we succeeded in bringing together two very different theoretical frameworks in a single computer model," says Professor Giraud. "On the other, we have shown that neuronal oscillations most likely rhythmically align the endogenous functioning of the brain with signals that come from outside via the sensory organs. If we put this back in predictive coding theory, it means that these oscillations probably allow the brain to make the right hypothesis at exactly the right

More information: Sevada Hovsepyan et al. Combining predictive coding and neural oscillations enables online syllable recognition in natural speech, *Nature Communications* (2020). DOI: 10.1038/s41467-020-16956-5

Provided by University of Geneva