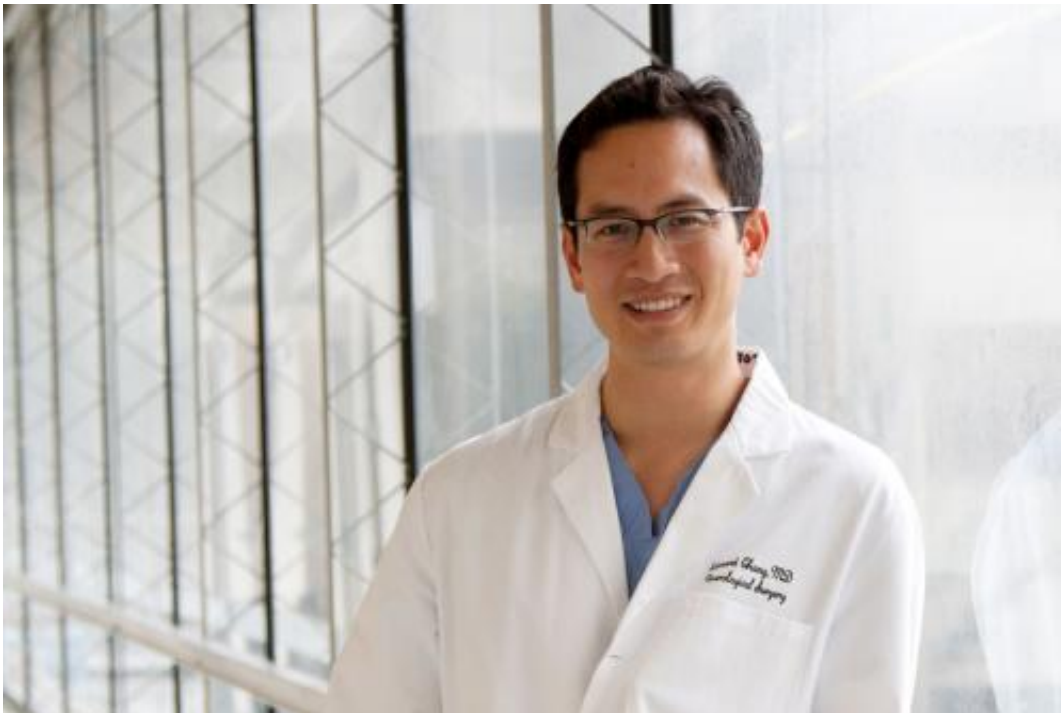# New study reveals how the brain recognizes speech sounds (w/ video)

January 30 2014

This is Edward Chang, M.D., senior author of the study. Credit: UC San Francisco

UC San Francisco researchers are reporting a detailed account of how speech sounds are identified by the human brain, offering an unprecedented insight into the basis of human language. The finding, they said, may add to our understanding of language disorders, including dyslexia.

Scientists have known for some time the location in the brain where speech sounds are interpreted, but little has been discovered about how this process works.

Now, in *Science Express* (January 30th, 2014), the fast-tracked online version of the journal *Science*, the UCSF team reports that the brain does not respond to the individual sound segments known as phonemes—such as the b sound in "boy"—but is instead exquisitely tuned to detect simpler elements, which are known to linguists as "features."

This organization may give listeners an important advantage in interpreting speech, the researchers said, since the articulation of phonemes varies considerably across speakers, and even in individual speakers over time.

The work may add to our understanding of reading disorders, in which printed words are imperfectly mapped onto speech sounds. But because speech and language are a defining human behavior, the findings are significant in their own right, said UCSF neurosurgeon and neuroscientist Edward F. Chang, MD, senior author of the new study.

"This is a very intriguing glimpse into speech processing," said Chang, associate professor of neurological surgery and physiology. "The brain regions where speech is processed in the brain had been identified, but no one has really known how that processing happens."

Although we usually find it effortless to understand other people when they speak, parsing the speech stream is an impressive perceptual feat. Speech is a highly complex and variable acoustic signal, and our ability to instantaneously break that signal down into individual phonemes and then build those segments back up into words, sentences and meaning is a remarkable capability.

Because of this complexity, previous studies have analyzed brain responses to just a few natural or synthesized speech sounds, but the new research employed spoken natural sentences containing the complete inventory of phonemes in the English language.

To capture the very rapid brain changes involved in processing speech, the UCSF scientists gathered their data from neural recording devices that were placed directly on the surface of the brains of six patients as part of their epilepsy surgery.

The patients listened to a collection of 500 unique English sentences spoken by 400 different people while the researchers recorded from a brain area called the superior temporal gyrus (STG; also known as Wernicke's area), which previous research has shown to be involved in speech perception. The utterances contained multiple instances of every English speech sound.

Many researchers have presumed that brain cells in the STG would respond to phonemes. But the researchers found instead that regions of the STG are tuned to respond to even more elemental acoustic features that reference the particular way that speech sounds are generated from the vocal tract. "These regions are spread out over the STG," said first author Nima Mesgarani, PhD, now an assistant professor of electrical engineering at Columbia University, who did the research as a postdoctoral fellow in Chang's laboratory. "As a result, when we hear someone talk, different areas in the brain 'light up' as we hear the stream of different speech elements."

"Features," as linguists use the term, are distinctive acoustic signatures created when speakers move the lips, tongue or vocal cords. For example, consonants such as p, t, k, b and d require speakers to use the lips or tongue to obstruct air flowing from the lungs. When this occlusion is released, there is a brief burst of air, which has led linguists

to categorize these sounds as "plosives." Others, such as s, z and v, are grouped together as "fricatives," because they only partially obstruct the airway, creating friction in the vocal tract.

The articulation of each plosive creates an acoustic pattern common to the entire class of these consonants, as does the turbulence created by fricatives. The Chang group found that particular regions of the STG are precisely tuned to robustly respond to these broad, shared features rather than to individual phonemes like b or z.

Chang said the arrangement the team discovered in the STG is reminiscent of feature detectors in the visual system for edges and shapes, which allow us to recognize objects, like bottles, no matter which perspective we view them from. Given the variability of speech across speakers and situations, it makes sense, said co-author Keith Johnson, PhD, professor of linguistics at the University of California, Berkeley, for the brain to employ this sort of feature-based algorithm to reliably identify phonemes.

"It's the conjunctions of responses in combination that give you the higher idea of a phoneme as a complete object," Chang said. "By studying all of the speech sounds in English, we found that the brain has a systematic organization for basic sound feature units, kind of like elements in the periodic table."

  More information: "Phonetic Feature Encoding in Human Superior Temporal Gyrus," by N. Mesgarani et al. *Science*, 2014.

Provided by University of California, San Francisco