Visual cues help people understand spoken words

HOUSTON -- (March 4, 2009) -- Understanding what a friend is saying in the hubbub of a noisy party can present a challenge – unless you can see the friend's face.

New research from Baylor College of Medicine in Houston and the City College of New York shows that the visual information you absorb when you see can improve your understanding of the spoken words by as much as sixfold.

Your brain uses the visual information derived from the person's face and lip movements to help you interpret what you hear, and this benefit increases when the sound quality rises to moderately noisy, said Dr. Wei Ji Ma, assistant professor of neuroscience at BCM and the report's lead author, in a report that appears online today in the open access journal PLoS ONE.

What you see affects what you hear

"Most people with normal hearing lip-read very well, even though they don't think so," said Ma. "At certain noise levels, lip-reading can increase word recognition performance from 10 to 60 percent correct."

However, when the environment is very noisy or when the voice you are trying to understand is very faint, lip-reading is difficult.

"We find that a minimum sound level is needed for lip-reading to be most effective," said Ma.

Lip-reading data

This research is the first to study word recognition in a natural setting, where people report freely what they believe is being said. Previous experiments only used limited lists of words for people to choose from.

The lip-reading data help scientists understand how the brain integrates two different kinds of stimuli to come to a conclusion.

Ma and his colleagues constructed a mathematical model that allowed them to predict how successful a person will be at integrating the visual and auditory information.

People actually combine the two stimuli close to optimally, Ma said. What they perceive depends on the reliability of the stimuli.

"Suppose you are a detective," he said. "You have two witnesses to a crime. One is very precise and believable. The other one is not as believable. You take information from both and weigh the believability of each in your determination of what happened."

Example 1

Congruent audio and video

Example 2

Congruent audio and video with -12dB noise

Example 3

Congruent audio and video ("cartoon")

Example 4

Congruent audio and video ("cartoon") with -12dB noise

In a way, lip-reading involves the same kind of integration of information in the brain, he said.

Visual information

In experiments, videos of individuals were shown in which a person said a word (see Example 1). If the person is presented normally, the visual information provides a great benefit when it is integrated with the auditory information, especially when there is moderate background noise (see Example 2). Surprisingly, if the person is just a "cartoon" that does not truly mouth the word, then the visual information is still helpful, though not as much (see Examples 3 and 4).

In another study, the person mouths one word but the audio projects another, and often the brain integrates the two stimuli into a totally different perceived word (see Examples 5 and 6).

"The mathematical model can predict how often the person will understand the word correctly in all these contexts," Ma said.

Others who took part in this research include Xiang Zhou, Lars A. Ross, John J. Foxe and Lucas C. Parra of The City College of New York in New York City.

The full report can be found at http://dx.plos.org/10.1371/journal.pone.0004638.

Example 5

Incongruent audio and video (video = care, audio = tear)

Example 6

Incongruent audio and video with -12dB noise (video = care, audio = tear)

Last modified: May 1, 2009