

Study reveals vision's role in vowel perception

March 14 2018



In a study based at Brown University, researchers found that the motion and configuration of a speaker's lips are key components of the information people gather when distinguishing vowels in speech. In this case the difference is the position of the lips when making the 'oo' sound, as in goose, in English and in French. Credit: Masapollo et. al.

For all talkers, except perhaps the very best ventriloquists, the production of speech is accompanied by visible facial movements. Because speech is more than just sound, researchers set out to ascertain the exact visual information people seek when distinguishing vowel

sounds.

"An important and highly debated issue in our field concerns what is it that we are attending to in speech—what's the object of perception?" said lead author Matthew Masapollo, who conducted the research as a postdoctoral scholar at Brown University and is now at Boston University. "Another question that's debated is whether speech processing is special and distinct from other kinds of auditory processing since it is not purely an acoustic signal."

Resolving these questions would improve the scientific understanding of how we perceive speech, Masapollo said. That, in turn, could apply to the design of more intelligible online avatars and physical robots, and could even improve computer recognition of human speech and enhance communication devices for the hearing impaired.

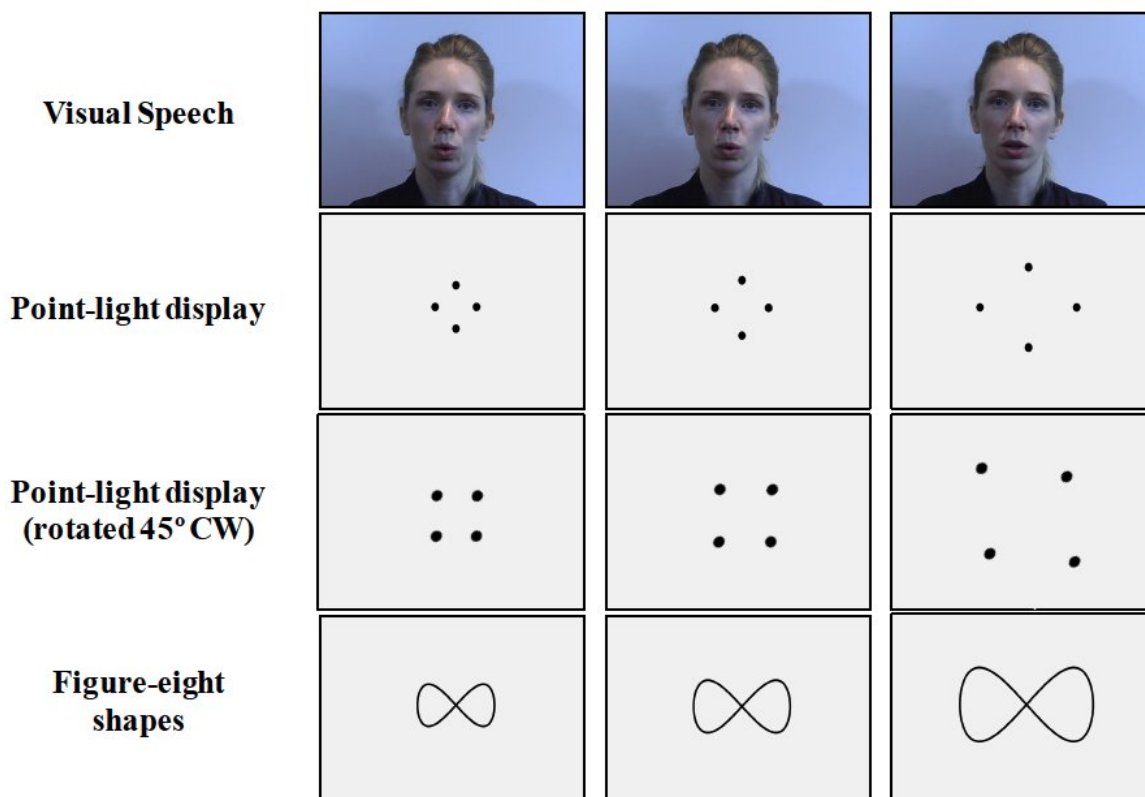
While scads of studies have investigated which audible features of speech are important, Masapollo said, far fewer have looked at which visual components are essential, despite evidence from phenomena as intuitive as lip reading that the sights of speech matter, too.

Through a series of experiments at Brown and McGill University in Montreal reported in the *Journal of Experimental Psychology: Human Perception and Performance*, Masapollo and colleagues found that when people perceive speech, they closely watch the form and motion of the lips. If either of those cues is missing, their ability to make subtle distinctions between [vowel](#) sounds suffers measurably.

"The findings demonstrate that adults are sensitive to the observable shape and movement patterns that occur when a person talks," said Masapollo, who did the work as a researcher in the lab of senior author James Morgan, a Brown professor of cognitive, linguistic and psychological sciences.

Exploiting differences in speech perception

Earlier this year, Masapollo set the table for the new study when he and co-authors Linda Polka and Lucie Ménard showed in the journal *Cognition* that people exhibit the same "directional asymmetry" in visually perceiving vowels that they do when hearing vowels: They are better at distinguishing between two versions of the "oo" sound, as in the word "loose," if the less extremely articulated version occurs first and then the more extreme version second. If the order is switched, they are much less likely to discriminate them—by sight or sound. While these directional effects may seem like a quirky instinct, they reflect a universal bias favoring vowels produced with extreme articulatory maneuvers. Current research is focused on uncovering what salient features or properties of extreme vowels give rise to these perceptual asymmetries.



Researchers employed many visual representations of lip motion to see which essential features really mattered. Masapollo et. al. Credit: Brown University

It turns out that this asymmetry plays out between French and English, being manifest in the bilingual speech of many Canadians. When speaking French, their articulation of "oo" is produced with more visible lip protrusion and tongue positioning than when making the same vowel sound in English.

For the new study, Masapollo realized that this asymmetry in vowel production and perception provided a great opportunity to determine which visual features matter in distinguishing subtle speech differences. He devised and led five experiments to ferret out exactly what visual information was pertinent to this asymmetry.

In the first, with help from Brown graduate student and co-author Lauren Franklin, he employed eye-tracking technology to measure where Brown student volunteers looked when watching videos of a bilingual Canadian woman make "oo" sounds in both French and English. Definitively, people watched the mouth, far more, for instance than the eyes.

But what about the mouth mattered? To determine if motion, rather than simply a particular position, was important, the next experiment presented students with a still frame rather than video. In experiment two, volunteers at McGill tried to distinguish "oo" speech using just still images of the same speaker. Without the cue of motion, the results showed, the asymmetry of French-English or English-French ordering no longer occurred, suggesting that motion is a key component in this

instinct of vowel perception.

In the next three experiments, the team continued to investigate which visual aspects of speech perception mattered among groups of Brown or McGill student volunteers. In experiment three, the subjects saw not a face, but an array of four dots in a diamond pattern that moved just like the speaker's lips did. When the speaker pursed her lips to make the "oo," the dots moved closer together, for example. Masapollo's hypothesis was that position and motion might matter together, even if the face isn't actually represented. In this experiment, people returned to showing the asymmetry suggesting that he was on the right track.

Experiment four was exactly the same but the dot pattern was rotated 45 degrees clockwise, showing more of a square than a diamond. Here the asymmetry didn't occur, suggesting that the orientation of the dots to represent a speech-making mouth matter. In experiment five, the motion was represented by a sideways figure eight that would move in a manner analogous to the speaker's lips. There, too, without even an essential form of a mouth, people didn't show their instinctual [asymmetry](#) of vowel perception. Mere motion, without the form and position of a mouth, was not enough.

"Overall, the picture that emerges is that perceptual asymmetries appear to be elicited by optical stimuli that depict both lip motion and configural information," the authors wrote.

To Masapollo, the results demonstrate that vision makes specific contributions to perceiving speech.

"The findings of the present research suggest that the information we are attending to in speech is multimodal, and perhaps gestural, in nature," Masapollo said. "Our perceptual system appears to treat auditory and visual [speech](#) information similarly."

More information: Matthew Masapollo et al, Asymmetries in unimodal visual vowel perception: The roles of oral-facial kinematics, orientation, and configuration., *Journal of Experimental Psychology: Human Perception and Performance* (2018). [DOI: 10.1037/xhp0000518](https://doi.org/10.1037/xhp0000518)

Provided by Brown University

Citation: Study reveals vision's role in vowel perception (2018, March 14) retrieved 1 May 2024 from <https://medicalxpress.com/news/2018-03-reveals-vision-role-vowel-perception.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--