



## UNDERSTANDING SPEECH WITH OUR EARS AND EYES

**Edin Šabić<sup>1</sup>, Michael C. Hout<sup>1,2\*</sup>, Justin A. MacDonald<sup>3</sup>, Daniel Henning<sup>3</sup>, Hunter Myüz<sup>3,4</sup> and Audrey Morrow<sup>3,5</sup>**

<sup>1</sup>Vision Sciences and Memory Laboratory, Department of Psychology, New Mexico State University, Las Cruces, NM, United States

<sup>2</sup>Addison Care Virtual Reality and Augmented Reality Laboratory, New Mexico State University, Las Cruces, NM, United States

<sup>3</sup>Hearing Enhancement and Augmented Reality Laboratory, Department of Psychology, New Mexico State University, Las Cruces, NM, United States

<sup>4</sup>Physiology-Sensing Intelligent Optimization Nucleus Laboratory, Flight Systems Branch, National Aeronautics and Space Administration, Houston, TX, United States

<sup>5</sup>Samaha Lab, Department of Psychology, University of California, Santa Cruz, Santa Cruz, CA, United States

### YOUNG REVIEWER:



ROMEO

AGE: 10

Understanding people when they are speaking seems to be an activity that we do only with our ears. Why, then, do we usually look at the face of the person we are listening to? Could it be that our eyes are also involved in understanding speech? We designed an experiment in which we asked people to try to comprehend speech in different listening conditions, such as someone speaking amid loud background noise. It turns out that we can use our eyes to help understand speech, especially when that speech is difficult to hear clearly. Looking at a person when they speak is helpful because their mouth and facial movements provide useful clues about what is being said. In this article, we explore how visual information influences how

**we understand speech and show that understanding speech can be the work of both the ears and the eyes!**

## **CAN A SPEAKER'S FACIAL MOVEMENTS HELP US HEAR?**

Have you ever wondered why people often look at the faces of those they are listening to? This behavior feels natural and automatic and, in many cultures, it is polite to make eye contact during conversation. But speech, like all other sounds, is made up of sound waves, which only the ears can pick up. So, what are the eyes doing while we listen? It turns out that understanding speech involves both the eyes and the ears, especially when listening conditions are not ideal, like when loud music is playing in the background. In non-ideal listening conditions, a listener might choose to look more closely at a speaker's face and mouth, to better understand what is being said.

Previous research on how the ears and eyes work together to help us understand speech has focused mostly on peoples' understanding of simple speech sounds, such as "ba" or "ga." These short component sounds of words are known as **phonemes**. When phonemes are combined, they can create all the words we might hear. Why did previous researchers use such simple sounds? One reason is that phonemes do not have any meaning, so listeners cannot make an educated guess as to what they heard by using other information spoken along with the sounds, but instead must listen very precisely to the phonemes. Testing phonemes one at a time is a straightforward approach, but when we are having conversations in our everyday lives, we do not perceive individual phonemes, but rather full, complex words and sentences. For this reason, our group conducted a study [1] in which we asked individuals to try to understand the sentence a speaker said during various listening conditions. Watching someone's facial expressions can, of course, provide us with information about the emotions the speaker may be feeling, but can facial movements also help us better hear what the speaker is saying? Our question was simple: Why might people look at the face of the speaker they are listening to during speech comprehension?

## **WHERE ARE THE EYES LOOKING?**

To understand how and why people use their eyes during listening, we needed a way to monitor eye movements. How can scientists track eye movement? We use devices called **eye trackers!** The type of eye tracker we used in this study has a camera that can tell where the listener is looking (Figure 1). The camera accomplishes this by detecting the center of the pupil—which is the black, central part of the eye—and also by using a sensor to keep track of how light is reflected by the eye. This light is produced by a non-moving light source next to the eye tracker camera. Light reflection information is

### **PHONEMES**

The smallest possible units of sound in spoken language. When phonemes are combined, they make up all the possible words in any given language.

### **EYE-TRACKER**

A computerized device that determines where a person is looking. Eye-trackers usually contain a camera and a light source that is used to detect how the viewer's eyes are rotated.

## Figure 1

**(A)** A research participant sitting in front of the eye-tracker. Participants sit with their faces in a chin rest while the eye tracker (located between the participant and the monitor) detects where they are looking. **(B)** A close-up of the eye tracking hardware, with the camera on the left and the infrared illuminator on the right, and a view of the computer that is monitoring the participant's eyes.



Figure 1

## FIXATION

A time period during which the eyes are almost completely still. Fixations are quite brief, often lasting only a fraction of a second.

## SACCADE

A very fast movement of both eyes that occurs in between fixations, allowing the eyes to quickly “jump” (rotate) to bring a new part of the world into focus.

used to figure out how the eye is rotated, and therefore to calculate where the listener is looking on the screen.

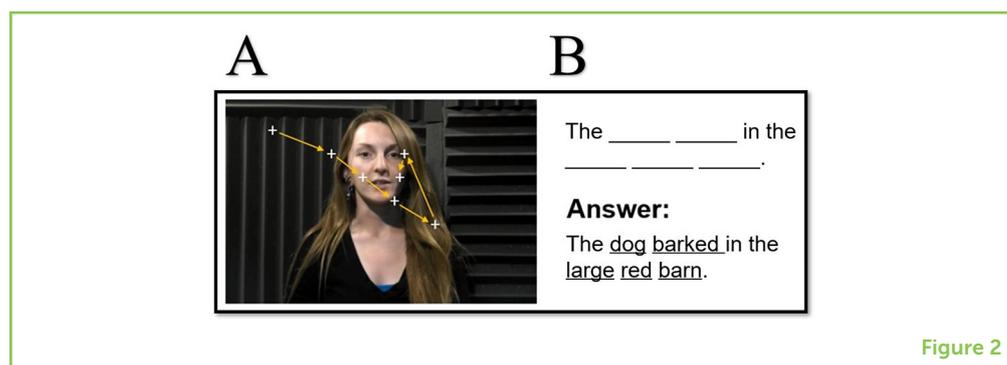
It is important to understand what type of information an eye tracker collects. One type of eye behavior is called **fixation**, which is when a person keeps their gaze focused on one location for a short period of time. For example, you might fixate on a camera lens while someone is taking your picture. Another type of eye behavior is a **saccade**, which is a quick movement (of both eyes in the same direction) to a new location, like when you quickly move your eyes away from the lens to look at something else. Saccades are very fast—they are one of the fastest movements produced by humans! When we are moving our eyes in daily life, we are using a combination of both fixations and saccades. An eye tracker can identify and categorize eye movements as either fixations or saccades, which is important so that researchers can better understand what our eyes do as we accomplish tasks.

## HOW DID WE STUDY THE ROLE OF THE EYES IN UNDERSTANDING SPEECH?

To understand the role of the eyes in listening, we designed a study in which participants completed a simple task: they listened to audio clips of people speaking or watched short movie clips, and then reported what they heard by completing fill-in-the-blank sentences (Figure 2). We used the eye tracker to record where people were looking while they were listening. For half of the participants, we also simulated hearing impairment by playing a loud static noise as they completed both tasks. This made it seem like half of our participants had some level of hearing loss, although they actually had no trouble hearing in ordinary circumstances. For both groups of participants, various levels of background noise were also played during each repeat of the task. The video clips started out easy to hear, with the speech being much louder than the background noise. Hearing became harder and harder as time went by, and at every sixth clip the background noise was just as loud as the speech! Then the process started over and proceeded for another set of six clips.

## Figure 2

(A) A sample movie clip of a woman speaking a sentence. Drawn on top of the figure, we see a sample pattern of fixations (crosses) and saccades (arrows) recorded by the eye tracker. (B) The fill-in-the-blank prompt that the participants answered after hearing the sentence (top), and the correct answer, which was not shown to the participants (bottom).



## WHAT DID OUR RESULTS SHOW?

In the audio-only task in which no videos were shown, the participants in the normal hearing condition found it easier to recognize what was said than did participants in the simulated hearing-impairment condition. This makes sense, because people with hearing impairments often have more difficulty understanding speech. However, this only informed us about how well people can understand speech when the speaker's face is not visible. How would these two groups perform when the participants could also see the speaker's face?

When we looked at the results of the experiments in which participants watched movies of the speaker while they listened and combined those results with the audio-only experiments, we found something interesting (Figure 3). We found that participants with normal hearing understood what the speaker said with the same accuracy, whether they saw the speaker's facial movements or not. In other words, they performed identically in the task across the audio and movie conditions. However, when we looked at the simulated-hearing-impairment group, an interesting finding emerged—when this group could see the speaker's face, they did a lot better on the accuracy test!

We also discovered a couple of interesting things that participants were doing with their eyes during the movies. We found that participants in the simulated hearing-impairment condition made fewer total fixations, which means they focused their eyes on fewer locations on the screen than the group with normal hearing did. We also found that, as the background noise became louder, participants' fixations became longer. That is, when participants in the simulated-hearing-impairment condition fixated upon something, they did so for a longer time than the group with normal hearing did. This suggests that in more difficult listening conditions, such as when hearing is impaired or there is a lot of background noise, people tend to fixate longer on areas of the face responsible for speech production. That means we may use clues from the speaker's face to better understand speech when listening is difficult. You probably do

### Figure 3

Our experiment found that the normal hearing group (dotted line) had an easier time understanding the sentences and did equally well in the audio-only condition and the movie condition. The simulated-hearing-impairment group (solid line) had a harder time but did much better in the movie condition when they could see the speaker's face! This tells us that our eyes can help us understand what a person is saying.

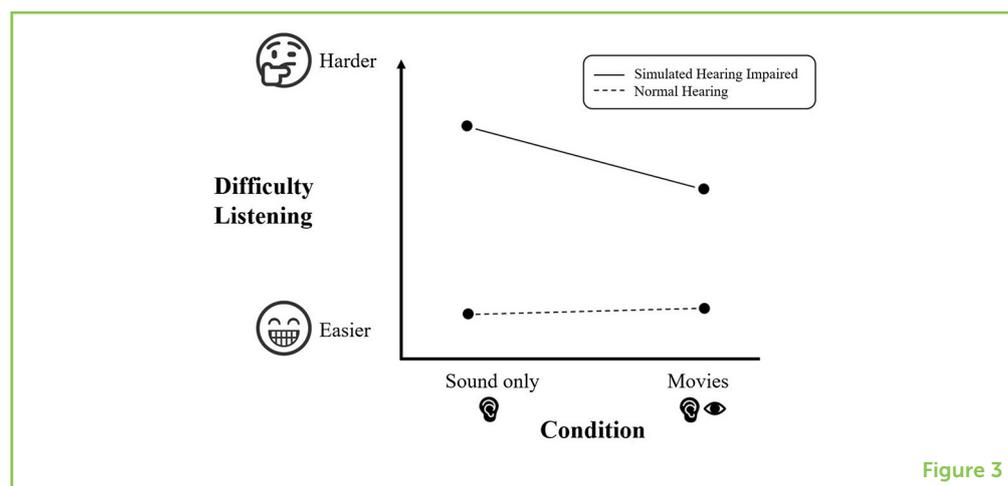


Figure 3

this, too! For instance, if you are at a loud party, you might find yourself watching your friend's mouth and face more closely to be sure that you understand what he or she is saying.

### WHAT DO OUR RESULTS MEAN?

Our experiment suggests that participants with simulated hearing impairment could make use of visual information when auditory information was more difficult to comprehend. This finding highlights how flexible our speech perception behavior can be, and how we can adapt by using our eyes to help us understand speech. Imagine someone you know who has trouble hearing, like a grandparent. Have you ever noticed that they tend to look at your face more often when you speak than people with unimpaired hearing do? This might be why! Your grandparent is probably using vision to understand you better.

So, let us take a step back to our research question and ask, "Why do people look at the face of the speaker they are listening to?" In the case of background noise, it appears that people may look at the speaker's face because there are facial cues that provide information about what the person is saying. Our research shows that our eyes can help us understand what is being said—particularly when it is difficult for us to understand what is said based on sound alone. Further, people are impressively able to adapt to understand speech by using their eyes to watch facial movements. It turns out that our eyes can be involved in speech perception too, and that our eyes and ears can work together in this task—even when we do not notice it! These findings not only help us to better understand the basic processes of speech perception, but might also help researchers better design technologies to support those with hearing impairment.

## ORIGINAL SOURCE ARTICLE

Šabić, E., Henning, D., Myuz, H., Morrow, A., Hout, M. C., and MacDonald, J. 2020. Examining the role of eye movements during conversational listening in noise. *Front. Psychol.* 11:200. doi: 10.3389/fpsyg.2020.00200

## REFERENCES

1. Šabić, E., Henning, D., Myuz, H., Morrow, A., Hout, M. C., and MacDonald, J. 2020. Examining the role of eye movements during conversational listening in noise. *Front. Psychol.* 11:200. doi: 10.3389/fpsyg.2020.00200

**SUBMITTED:** 04 June 2020; **ACCEPTED:** 20 May 2021;

**PUBLISHED ONLINE:** 28 June 2021.

**EDITED BY:** Rich Ivry, University of California, Berkeley, United States

**CITATION:** Šabić E, Hout MC, MacDonald JA, Henning D, Myüz H and Morrow A (2021) Understanding Speech With Our Ears and Eyes. *Front. Young Minds* 9:569624. doi: 10.3389/frym.2021.569624

**CONFLICT OF INTEREST:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**COPYRIGHT** © 2021 Šabić, Hout, MacDonald, Henning, Myüz and Morrow. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## YOUNG REVIEWER

### ROMEO, AGE: 10

Romeo loves to play a card game called magic the gathering and many other games. His birthday is October 13 and he is 10 years old. His favorite book and movie is "the hunger games." Romeo has a very short attention span and it is kicking in, so BYE.

## AUTHORS

### EDIN ŠABIĆ

Edin is a recent graduate of New Mexico State University, where he earned his doctorate degree in Experimental Psychology. His research focuses on how people interact with technologies toward the aim of improving these experiences. Outside of research, Edin enjoys hiking, playing soccer, and traveling to new places.



**MICHAEL C. HOUT**

Michael C. Hout is an Associate Professor in the Department of Psychology at New Mexico State University, and a Program Director at the National Science Foundation. His research focuses primarily on visual cognition (including search, attention, eye movements, and memory) and the development of methods of collecting similarity data for use in multi-dimensional scaling. He has won several awards for research and teaching, including the Rising Star award from the Association for Psychological Science. In his limited free time, he enjoys walking his dog, running, hiking, and playing hockey. \*mhout@nmsu.edu

**JUSTIN A. MACDONALD**

Justin received his doctorate in Quantitative-Mathematical Psychology from Purdue University in 2003. He is currently an Associate Professor in the Department of Psychology at New Mexico State University. His research interests include visual and auditory perception, and statistics.

**DANIEL HENNING**

Daniel is a doctoral student in Engineering Psychology at New Mexico State University. His research focuses on auditory perception and how to improve products and technology for a better user experience. Daniel enjoys hiking and playing the drums.

**HUNTER MYÜZ**

Hunter received her doctorate in Experimental Psychology in 2020. She is currently a User Experience researcher in the technology industry. Her research ranges from social topics, such as religion, to practical applications of statistics and design. Hunter enjoys many things, including: animals, crafts, and playing music and sports.

**AUDREY MORROW**

Audrey received her Bachelor's degree from Towson University in Maryland and her Master's degree from New Mexico State University, focused in cognitive psychology. Her thesis research used brain stimulation to try to reduce the negative effects that stress has on cognitive performance. Currently, Audrey is a Ph.D. student in the psychology department at University of California, Santa Cruz where she uses electroencephalogram (EEG) to study brain waves during different attentional and perceptual states. Audrey enjoys arts and crafts, hiking, and the occasional board game.