Study suggests we can recognize speakers only from how faces move when talking

April 18 2018, by Alexandra Jesse

Alexandra Jesse, cognitive psychologist at UMass Amherst, says her findings may have practical implications for facial recognition software. It may be made more reliable if it used a person speaking a short sentence rather than a static photo, because speech, with its combination of static plus dynamic features, offers a more complex personal identifier and might be harder to falsify. Credit: UMass Amherst

Results of a new study by cognitive psychologist and speech scientist Alexandra Jesse and her linguistics undergraduate student Michael Bartoli at the University of Massachusetts Amherst should help to settle a long-standing disagreement among cognitive psychologists about the information we use to recognize people speaking to us.

She explains, "It is of great importance to us in our lives to be able to recognize a person we met before, when we meet them again later. People who study face perception have been arguing that when we meet new people, all we are learning about their faces are so-called static features that don't change, like the shape and size of their face and skin color. They dismiss the idea that we also learn to recognize a person by the unique way in which their mouth and facial muscles moves as they talk."

"But people in my field of speech perception have a strong sense that these dynamic features are important," Jesse adds. "We know that when people can see each other in conversations, they recognize speech not just from listening alone but also from lip-reading. In my lab, we study this audio-visual speech perception. The missing link, and the reason for this study, was to show that listeners can use visual dynamic features to learn to recognize who is talking."

In this study, the UMass Amherst researchers found in a series of experiments that adult listeners can indeed learn to recognize previously unfamiliar speakers from seeing only the motion they produce while talking. Listeners learned to recognize the personal motion "signatures" of new speakers readily and from limited exposure. Results are online now in the journal *Cognition* and are expected to appear in the July print edition.

Jesse says, "In all of our experiments, we found learning with a very brief exposure to the previously unfamiliar speakers. Most people had

learned from seeing each speaker less than eight times. Not only do we learn very quickly but we're also not simply learning to recognize a speaker from how they said a particular sentence, rather we are learning to recognize that speaker from any sentence. It's a very quick, efficient process. It demonstrates that we use speaking-related motion not just to recognize what the person is saying but also to recognize the individual."

Their findings may have important practical implications for personal and facial recognition technologies and software such as are used in airport security, for example, the researcher says. Such technologies may be made more reliable if they used a person speaking a short sentence rather than a static photo, because speech, with its combination of static features plus dynamics, offers more data, a more complex personal identifier. "It might be a lot harder to falsify," she adds.

For this investigation into whether people learn to recognize a person from facial motion as they talk, with no other cues, the researchers generated what they call configuration-normalized point-light displays of faces that show only the biological motion that speakers produce while saying short sentences.

To create point-light displays, they glued 23 white paper dots on the faces of two different speakers and videotaped them as they spoke simple sentences. The researchers showed videos of these speakers to listeners with only the dots seen moving against a black background, no sound and no facial details.

In the initial training phase of the experiment, listeners watched each video before responding with one of two names to indicate who they thought they saw. At first, they had to guess, as they did not know these speakers, but through feedback on their performance, they learned. At a subsequent test phase, no feedback was given, and listeners saw also videos of the speakers saying new sentences.

"Listeners learned to identify two speakers, and four speakers in another experiment, from visual dynamic information alone. Learning was evident already after very little exposure," the authors report. Further, they says, listeners formed abstract representations of visual dynamic signatures that allowed them to recognize speakers" even when seeing them speak a new [sentence](#).

Jesse points out, "Speech perception is a really difficult task - and recognizing who the speaker is can help with it. One thing this research shows is that we're not done as adults with learning, we are constantly learning about the new speakers we meet. As we get older, it becomes more difficult to recognize from listening alone what a speaker says and who they are, as does recognizing faces from static features."

She adds, "We already know that as we get older, seeing a speaker becomes more important for recognizing what they are saying. Based on our study though, we think that seeing a [speaker](#) may also become more important for recognizing who is talking to us, which then may have an indirect effect on [speech perception](#), as well."

 More information: Alexandra Jesse et al, Learning to recognize unfamiliar talkers: Listeners rapidly form representations of facial dynamic signatures, *Cognition* (2018). [DOI: 10.1016/j.cognition.2018.03.018](#)

Provided by University of Massachusetts Amherst