

Science Education Collection

The McGurk Effect

URL: <http://www.jove.com/science-education/10295>

Overview

Source: Laboratory of Jonathan Flombaum—Johns Hopkins University

Spoken language, a singular human achievement, relies heavily on specialized perceptual mechanisms. One important feature of language perception mechanisms is that they simultaneously rely on auditory and visual information. This makes sense, because until modern times, a person could expect that most language would be heard in face-to-face interactions. And because producing specific speech sounds requires precise articulation, the mouth can supply good visual information about what someone is saying. In fact, with an up-close and unobstructed view of someone's face, the mouth can often supply better visual signals than speech supplies auditory signals. The result is that the human brain favors visual input, and uses it to disambiguate inherent ambiguity in spoken language.

This reliance on visual input to interpret sound was described by Harry McGurk and John Macdonald in a paper in 1976 called *Hearing lips and seeing voices*.¹ In that paper, they described an illusion that arises through a mismatch between a sound recording and a video recording. That illusion has become known as the McGurk effect. This video will demonstrate how to produce and interpret the McGurk effect.

Procedure

1. Stimuli

1. To make McGurk effect stimuli you will need a video camera-the kind on a smartphone is fine.
2. You will also need a computer to control the presentation of the videos to a naïve subject.
3. Point your camera at yourself, so that your head fills the display.
4. Make four recordings. Each should be 10 s long. In each of the four recordings, you'll repeat a word 10 times, about 1/s. Here are the words: bane, gain, pan, can. Try to say the words in each video at a similar pace.

2. Inducing the Illusion

1. To induce the illusion, you could splice together the sound from one video and the picture from another. But that's not really necessary. It is easier to just do it by using your phone and a computer simultaneously. Here is how.
2. On your computer desktop open up the video in which you are saying gain. Turn off the sound, and play the video.
3. On your phone open up the video in which you are saying bane. Put the phone behind the computer screen so that the sound can be heard, but the video can't be seen. Play the video.
4. Ask an observer to watch the computer screen while listening, and when the video is done playing, ask them what they heard.
5. Do the same for the pan/can videos: Play the picture stream of you saying can while your phone plays the audio stream from the pan video. Ask the participant what she heard.

Results

Remember, the sounds played to your observer are either the words bane or pan. But in the accompanying videos, the words being articulated are gain and can respectively. So which words will people actually hear? The answer is most often none of those four. Instead, the typical result is that observers in the bane/gain condition will hear the word Dane. And observers in the pan/can condition will hear the word tan.

To understand why we need to understand a little bit about how phonemes are produced. A phoneme is a minimal unit of speech sound. The words bane and gain have the same phonemes in all positions but the first. In the word bane the first phoneme is a b sound, denoted /b/. In the word gain it is the sound /g/. The remaining sounds are the same-which is why the words rhyme. **Figure 1** breaks down the McGurk effect in terms of the initial phonemes in these examples. When /b/ is shown and /g/ is played, people hear /d/. The word Dane in other words also rhymes with bane and gain, with a one phoneme difference right at the beginning.

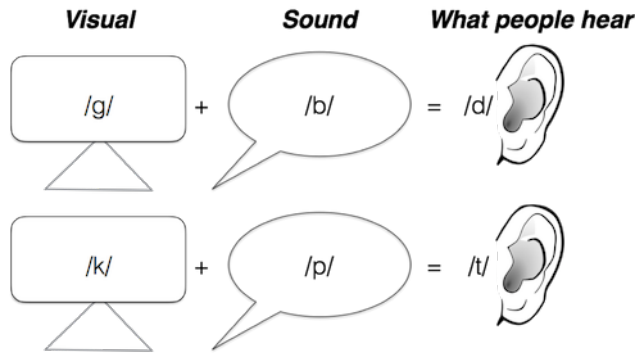


Figure 1: The McGurk effect happens when there is a mismatch between a phoneme that is articulated in a visual presentation and different phoneme is played simultaneously through speakers. With phonemes that share certain articulation properties, the result heard may not match either of the mismatching stimuli. In the mismatch causes a third sound to be heard. Specifically, a visual /g/ with an auditory /b/ causes the phoneme /d/ to be heard. This is why a visual gain with an auditory bane results in Dane being heard. Similarly, a visual /k/ with an auditory /p/ leads the sound /t/ to be heard. That's why can/pan produces tan in the McGurk effect.

Why do conflicting /b/ and /g/ produce a /d/ specifically? Well, /b/, /g/, and /d/ are really not that different from one another, especially in terms of how they are produced. The three basically involve moving the same amount of air from a person's larynx through their mouth, with just a difference in where the speaker places a small obstruction. When someone makes a /b/ sound, they use their lips to obstruct the air; this is known as a labial point of articulation. For a /g/ sound, the point of articulation is palatal—it is far in the back of the mouth. And for a /d/ sound, the point of articulation is known as dental because people obstruct airflow through the mouth by touching their tongues to the top teeth. **Figure 2** shows the relative points of articulation for the six phonemes in the McGurk effect.

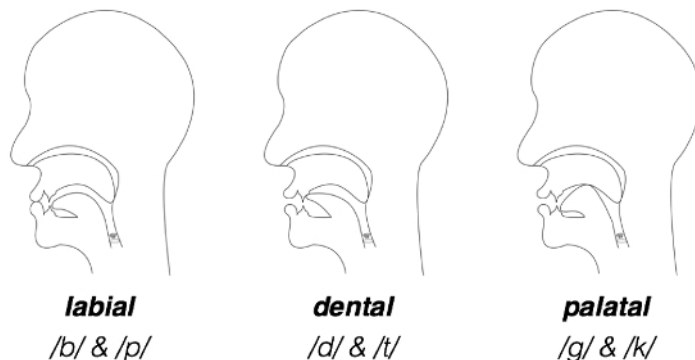


Figure 2: Humans produce sounds by moving air through their throats and mouth. This involves vibrations in the larynx. A given set of vibrations produced in the larynx can produce multiple different phonemes by obstructing the flow of air. The place where an obstruction is placed to create a specific sound is called the point of articulation. Three important points of articulation are known as labial, referring to the lips; dental, referring to the teeth; and palatal, referring to the palate, or the back roof of the mouth. The figure shows how the phonemes produced and heard in the McGurk effect differ in terms of their points of articulation.

Now that you know a bit about how these sounds are produced, the logic of the McGurk effect should be more apparent. It works like this: Your brain knows that some phonemes are actually pretty similar to one another. In the McGurk effect the word bane is played to the observer, led off by a /b/ sound. But the face in the video is moving their mouth as they would make a /g/ sound, and the word gain. The brain therefore receives conflicting inputs from the eyes and ears. To resolve the conflict, the brain comes to the conclusion that the truth is probably someplace in between. Since /d/ is the sound between /b/ and /g/-in terms of production—that's what people hear. The same explanation applies for turning the conflict between pan and can into tan. /p/ is a labial sound, and /k/ is a palatal sound. The dental one in between is /t/.

Applications and Summary

One place that the McGurk effect has been important is in understanding how very young infants learn spoken language. A study in 1997 was able to show that even 5-month-old infants perceive the McGurk effect.² This is important because it suggests that visual information may be used by infants to solve a major challenge to learning language—parsing a continuous audio stream into its units. Think about how a foreign language spoken at its normal speed can seem like such a jumble that you might not even know where the word boundaries are. Well, if all languages are foreign to infants, then how do they figure out where the words are? The McGurk effect suggests that they can rely on facial articulation patterns.

References

- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

2. Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, 59(3), 347-357.