How does the Human Brain perceive Speech?

Kate Watkins

Abstract: *Many animals can communicate but communication by speech is an exclusively human activity. Even computers struggle to recognise limited spoken instructions, making errors that young children do not make. Understanding speech is an obviously difficult task for a machine yet it is something we all do effortlessly. In this article, I describe some of the research being done to understand how the human brain accomplishes this impressive feat.*

To understand how our brains perceive speech, we need first to consider how speech is created. When we speak, we produce a stream of air from the lungs that causes noise by vibration, much like the air released from the stretched neck of a balloon. The air squeezes through the glottis – a pair of muscles in the throat commonly known as the vocal cords, causing them to vibrate. This creates the buzz, which in turn forms the "voice". Simple changes in the tension and position of your vocal cords allow you to alter the pitch of the voice by altering the frequency of the vibration. Above the vocal cords, the buzzing stream of air passes through the throat, mouth, lips and nasal cavity. Movements of the tongue, lips and palate alter the shape of these cavities, which changes the resonant frequencies in the sound wave, just like when you blow across the mouth of a bottle and produce different sounds depending on the level of liquid. The differences in size and shape of the chambers in the vocal tract results in the production of speech sounds that we call vowels and consonants. Vowels are produced when air travels through the vocal tract unimpeded – try saying "ahhh" - whereas rapidly stopping and releasing air or squeezing it past the tongue and teeth produces consonants. Think of where you block the vocal tract when you say "patty-cake", for example.

The human vocal tract is specially adapted for speech. Our tongues are proportionally shorter and rounder than those of other primates and our larynx is lower in the throat. This has resulted in some costs – crowding of teeth in the shorter jaw (which we can thank for impacted wisdom teeth!) and a greater risk of choking by inhaling food. Other animals are capable of sophisticated vocalisations - for example, parrots and large sea mammals, such as whales[1] – but both species produce sound very differently to the way human speech is produced. There is one group of mammals with a vocal tract similar to ours – the pinnipeds (seals, sea lions and walruses). Anecdotal accounts of an orphaned harbour seal called "Hoover" describe how he produced speech-like imitations of humans[2]. But simply owning a vocal tract that resembles that of a human does not allow animals to produce sophisticated speech like humans; to do this, it appears that you require a human brain.

So how do we perceive speech? The continuous sound waves produced by the vocal tract during speech are highly variable yet remain understandable as speech. We can represent these sound waves in terms of changes in frequency and amplitude over time. Even so, these abstract representations show very few commonalities across speaker or even for words spoken by the same

speaker in different contexts. One cannot simply look at a waveform and identify a particular speech token with any degree of certainty. Speech recognition appears to be a hard problem, therefore, due to the lack of consistency in the acoustic signal, yet the brain can robustly perceive speech sounds regardless of the inconsistency.

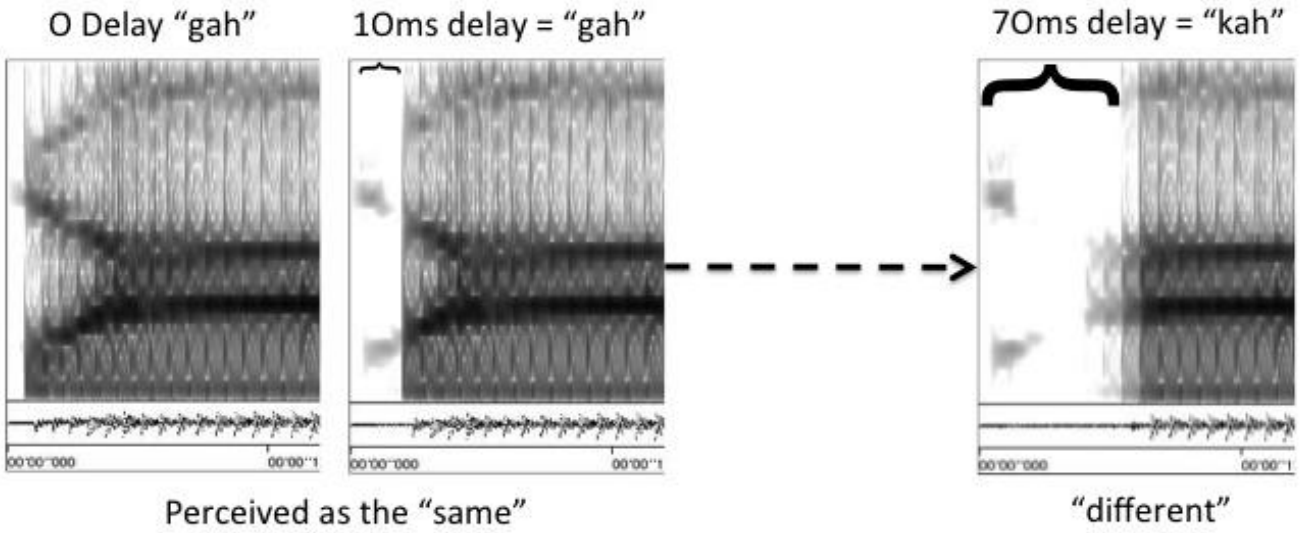

Figure 1. Picture of categorical perception: A spectrogram of the speech sound "gah" is shown on the left with the vocal tract release and the vocal cord vibration simultaneous. When we delay the vocal cord vibration by 10 ms, we still perceive the "gah" sound but when the delay exceeds about 40 ms, we start to hear a different sound "kah".

In the laboratory, we can demonstrate that the human brain perceives a whole range of acoustic signals as the ***same*** speech sound using computer-generated speech with software that morphs one speech sound into another. Take the "buh" sound at the beginning of a word like "bat", for example. This sound is produced by closing the lips briefly, allowing air pressure to build up behind the lips and then releasing the air through the lips while simultaneously vibrating the vocal cords. We can change the word "bat" into the word "pat" by simply increasing the delay between the release of air at the lips and the onset of the vibration of the vocal cords. Once this delay reaches about 40 milliseconds, most people's perception of the word will have changed from "bat" to "pat". What is interesting, however, is that perception of "bat" is unaffected as long as the delay remains under 40 milliseconds. If we ask people to say whether a version of "bat" with no delay or another version with a 30-millisecond delay sounds the same or different, they will confidently say they sound the same. Yet if we ask them whether one sound with a 30-millisecond delay and another with a 50-millisecond delay are the same or different, they will confidently say they sound different! It is as if there is a boundary at 40-milliseconds and everything on one side of the boundary is perceived as a "buh" sound and everything on the other side of the boundary is perceived as a "puh" sound. In psychology, we call this kind of perception, categorical perception – because the perceived element lies in one category of another. We perceive speech categorically but we perceive small differences in other non-speech sounds, such as a scale of piano tones, as continuous changes.

The categorical nature of speech perception led some scientists to conclude, "speech is special". If the same speech sound can be perceived from such a variety of

different acoustic representations, what could the brain use as its template to map these different sounds onto? One possibility is that speech perception is achieved by referencing it to speech production. This is the "motor theory of speech perception" and is one of the most controversial theories in psychology. Not least because it would predict that animals do not perceive speech sounds categorically and that people incapable of speaking (including infants) could not understand speech. Both of these predictions are incorrect. Nevertheless, the idea that the brain analyses speech by reference to the way it is produced has some appeal. It means that the same system is used for articulation and perception and it sidesteps the problem of needing an infinite number of representations of speech sounds to account for the massive variability in the acoustic signal.

A simple experiment demonstrates that our knowledge of the way speech is produced can affect the way it is perceived through a very robust illusion called the "McGurk effect". In this illusion, an audio recording of a speaker saying "bah" is played simultaneously with a video showing the speaker saying "gah".[3] What we perceive when watching this video is the sound "dah", which is actually neither seen nor heard. The brain appears to have integrated the conflicting information it has received through the ears (of a sound produced with lips closed) and through the eyes (of lips clearly open) and generated an illusory percept of a speech sound that lies somewhere in between these two. People usually only perceive just "bah" if they listen with their eyes closed or guess that "gah" is being spoken if the audio is muted. The McGurk illusion shows us that our brain makes use of its knowledge of how speech is produced in speech perception.

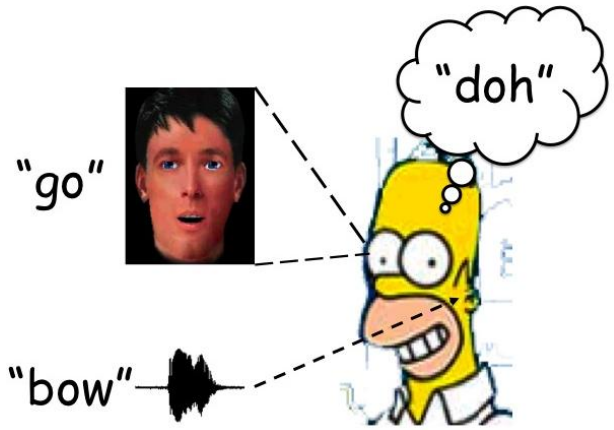

Figure 2: Picture on McGurk effect: Homer hears the word "bow" but sees the talking head saying the word "go", so he gets an illusory percept of the word "doh". It may not be in your vocabulary but I think Homer knows it!

In my laboratory, we are interested in understanding how far the system that produces speech – the motor system – contributes to speech perception. We use painless non-invasive brain stimulation techniques to temporarily interfere with the parts of the brain involved in speech production and see how this interference affects speech perception. The technique used is called transcranial magnetic stimulation or TMS. To stimulate the brain, we use a small coil, through which we briefly pass an electric current. The current generates a large but brief magnetic field around the coil. This change in the magnetic field induces current in anything capable of conducting electricity. Therefore, if we place the coil on the scalp, we can induce currents in the neurons near the surface of the brain. When we stimulate neurons in the motor cortex of the brain like this, we can elicit twitches in different muscles depending on where we stimulate. When we see a muscle twitch in the lips, we know we have found one part of the brain involved in speech production. We target this area with magnetic pulses to temporarily interfere with its function. The research volunteer is unaware of this impairment in function – if

we asked them to speak for example, they would still be able to speak. But with experimental tasks, we can show that this mild temporary interference of the speech motor cortex slightly impairs the volunteer's perception of speech sounds. They no longer hear the differences between similar sounds such as "bah" and "dah" as clearly as they did before the stimulation of the speech motor cortex. This very small effect lasts for about 20 minutes (and then their speech perception abilities return to normal). This experiment and others we are doing indicate that the parts of the brain involved in speech production are contributing directly to speech perception.

Neuroscience research continues to address questions about the remarkable ability of the human brain to communicate using speech. Improving knowledge in this area may further our understanding of speech impairment in children with developmental disorders and adults who acquire speech impairment through brain injury.

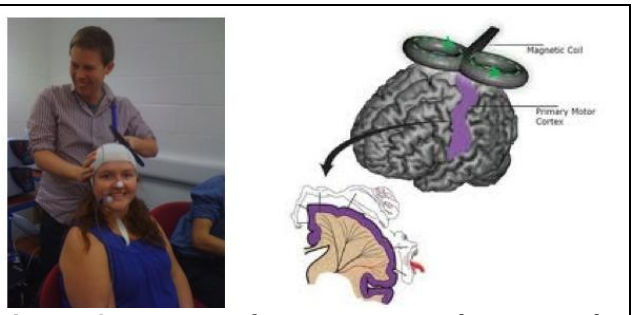

Figure 3: Picture of magnetic stimulation: Emily Connally (St. Anne's graduate student) is being stimulated using a magnetic coil by Dr. Jack Rogers (see how much they both enjoy this). The brain cartoon shows in purple the motor cortex, which when stimulated causes a muscle to twitch.
Photo Courtesy: Kate Watkins

References

[1]Ridgway S, Carder D, Jeffries M and Todd M (2012) Spontaneous human speech mimicry by a cetacean. Current Biology 22(20) R860-R861

[2]http://www.neaq.org/animals_and_exhibits/exhibits/individual_exhibits/harbor_seals_exhibit/hoover.php

[3]http://www.youtube.com/watch?v=aFPtc8BVdJk