Tracking the dance of the tongue using ultrasound imaging

A new project at the universities of Manchester and Lancaster promises to change how we think about vowel sounds. Sam Kirkham and Patrycja Strycharczuk explain how ultrasound imaging is used in linguistic research

Speaking involves the coordination of over one hundred muscles. The muscles work in tandem to move our tongue, lips and larynx. These movements cause the air outside of the vocal tract to vibrate, which listeners perceive as sounds. It is this incredible dance of the vocal organs that allows us to communicate thoughts from our mind into the mind of another person.

One of the challenging aspects of studying speech, however, is that most of the vocal tract is not directly visible to us. In other words, when we hear somebody speak, we know that their tongue is moving, but we often can't see it! As researchers in phonetics and speech science, we are often left to infer how the vocal tract is moving based on the audible speech signal. However, recent developments allow us to observe the movements of the tongue during speech, giving us access to new insights into speech communication.

What is ultrasound imaging?

Sound is the result of air molecules vibrating in the air around us. The rate at which those vibrations repeat is known as frequency and we typically measure frequency in terms of the number of cycles per second or Hertz (Hz). The average young healthy adult can hear sounds with frequencies between 20 Hz and 20,000 Hz, which we sometimes refer to as the acoustic range. Infrasound refers to sounds below this range of hearing (e.g. below 20 Hz), whereas ultrasound refers to frequencies that are higher than the upper limit of human hearing (e.g. above 20,000 Hz).

Other species can hear these very high frequency sounds: cats can hear up to around 60,000 Hz, whereas some bat species can hear up to 200,000 Hz. For example, bats cannot see particularly well, so they navigate their environment by emitting very high frequency sound waves and they listen to how these sound waves reflect off objects, allowing them to know where things are. Medical ultrasound imaging works in a similar way: we transmit very high frequency sounds at an object and record how quickly the sound bounces off that object. The most well-known kind of ultrasound imaging is the pregnancy scan, which involves using ultrasound to image a human fetus in the uterus. Medical ultrasound of this kind typically uses ultrasound of between two and ten Megahertz (MHz). One Megahertz is 1,000,000 Hz: in other words, very high frequency sound!

How is it used in linguistics?

Ultrasound in linguistics is commonly used for imaging the tongue. A probe is placed under the chin, which emits high-frequency sound waves that travel through the chin's tissue and reflect off the surface of the tongue. This reflection is recorded by the probe and used to construct an image, like the one in Figure 1. The bright white line is the surface of the tongue when viewed from the side of a person's face.

Ultrasound videos show lots of these snapshots over time, allowing us to obtain a dynamic picture of how the tongue is moving during speech. Ultrasound has been used for imaging the tongue during speech since the 1960s, and has become much more common for linguistic research in the subsequent decades. The biggest advantage of ultrasound is that it is significantly cheaper, more portable and less invasive than other imaging technologies, such as Magnetic Resonance Imaging.

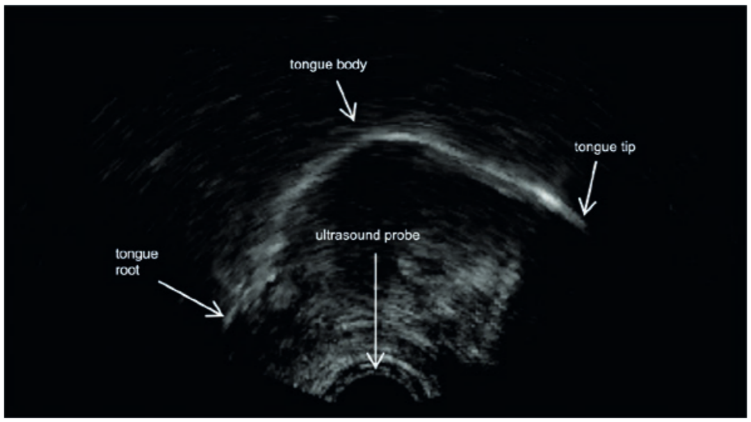

Figure 1

Tongue movement in English laterals

Imaging the movement of the tongue not only allows scientists to learn how speech works, but also how production of sounds differs between individuals and between accents, sometimes driving systematic change in language. The relationship between these phenomena can be quite complex. On the one hand, we all move our articulators differently, because we all have different anatomies. On the other hand, some aspects of articulatory movement become conventionalised within specific social groups, which is another way of saying that people speaking the same accent do something similar with their tongues.

In a series of studies, Patrycja Strycharczuk and James Scobbie explore these issues related to the production of the consonant spelled as 'l' in Southern British English. Despite the uniform spelling, this consonant sounds differently at the beginning of words, as in 'loop', and at the end of words, as in 'pool'. But that's not the end of it: the 'l' in 'pooling' can be slightly different from both the initial and the final 'l'. What's more, all these 'l's also differ between accents. The most noticeable manifestation of this is the vowel-like pronunciation that makes 'pool' sound like 'poow'.

Ultrasound research reveals additional complexities related to variation in 'l'. Some speakers produce their word-final 'l', also known as 'dark l', by pulling the tongue body backwards, followed by a forward tongue movement accompanied by raising of the tongue tip (see Figure 2 on page 10). Other speakers raise the tongue body as well as moving it backwards. This type of difference does not seem to map onto differences in sound, so they are an example of covert variation.

However, alongside covert variation, more systematic structural patterns also emerge. For instance, regardless of whether speakers retract or raise their tongue to make their final dark 'l's, they do it to a greater extent at the end of the word when the next word begins in a

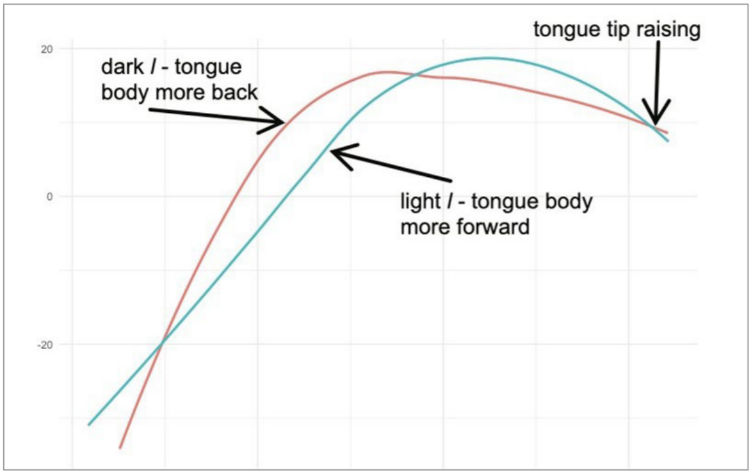

Figure 2

consonant (as in 'fool him') than if the next sound is a vowel (as in 'fool it'). This difference is also accompanied by a relative delay of the tongue tip raising.

Interestingly, the time interval between retraction and raising varies with age: younger speakers in the South of England have relatively darker 'l's with delayed tongue tip raising compared to their parents' generation. This difference cannot be explained by biological factors such as how fast we speak – young people generally speak faster, so we'd expect *less* delay, not more. It's also not because young people are lazy. In fact, it seems that young people go the extra mile retracting their tongue body, and then they still raise the tongue tip. Rather, the relative sequencing of the different tongue movements is a part of systematic sound change, or accent evolution, in which the nature of the sound changes systematically while no one (except a few linguists maybe) notices. The vocalic sounding 'w'-like 'l' is taking this process to its logical conclusion, and what's most surprising of all (this type of 'all' can still include tongue tip raising) is that while we might not be able to hear it, thanks to ultrasound, we can see that it's there.

What we see is not always what we hear

One interesting result of ultrasound research is our improved understanding of the relationship between tongue movements and the sound waves that propagate from our mouths, otherwise known as the acoustic signal. Sam Kirkham and Claire Nance investigated the production of vowels in Twi, an Akan language of Ghana. They also analysed the same speakers producing a set of similar vowels in Ghanaian English, something that was possible because all of the speakers in the study were bilingual (or, in many cases, multilingual). The aim of this study was to explore the relationship between tongue movements and the acoustic signal, and to investigate whether this varies between the two languages of a speaker.

An unexpected outcome of the study was that pairs of different vowels in Twi were produced using the same tongue position. The same speakers did not show this behaviour in their English; in other words, when two vowels sound different in Ghanaian English, they are produced using different tongue shapes. At first, this may seem a complex puzzle to solve: how can two sounds that involve the same tongue position sound different? Understanding this phenomenon requires us to think about the three-dimensional nature of the vocal tract. Ultrasound is typically used to image the tongue midsagittally, meaning that we only view a two-dimensional image of the tongue as viewed from a side profile. This means that we neglect any aspects of the vocal tract that occur in another plane.

In the case of the Twi vowels, previous studies of Akan languages have shown that the contrast between many pairs of vowels involves sideways expansion of the pharyngeal cavity, the space at the back of the throat. This would not be captured on the ultrasound image, as it only shows the tongue from one perspective and does not typically capture three-dimensional behaviour. Far from being a disappointing result, this finding brings us closer to understanding the role of the tongue in speech production – we are used to thinking about vowel contrasts in terms of tongue shape, but speakers use their vocal tracts in more complex ways. Kirkham and Nance use this knowledge to explain the discrepancy between tongue position and the acoustic signal in their results. This leads them to posit the existence of different articulatory goals for similar vowel sounds in a bilingual's two languages.

TARDIS: Tracking the tongue's movement across time

In our next research project, nicknamed TARDIS, we are combining ultrasound with magnetic tracking of sensors glued to the tongue to study the relationship between individual variation and accent variation when it comes to vowels. Vowels are at the heart of accent variation in English – the vowel in a word like 'face' sounds very different depending on whether you are in the South of England, North of England, Scotland or New Zealand. The same is true of many other vowels (but particularly diphthong vowels, which involve large inherent changes in quality), which can vary quite dramatically (think of words like 'price', 'goat' or 'choice' and how you'd pronounce them in different accents). These vowels are characterised by considerable movement of the articulators, so in this project, we are putting tongue movement front and centre. We will study such movement and develop a model of tongue movement that can accommodate individual variation while also capturing group-level regularities.

Applications of ultrasound tongue imaging

The issue of variation, and of what constrains it, is important when it comes to speech disorders. We have already seen that everyone moves their articulators in a slightly idiosyncratic way. Some of it doesn't matter (as in covert variation), but some of it does (as in accent variation). Variation is completely normal, and listeners can deal with quite a lot of it, but beyond certain ranges of variation, speech may become difficult to understand, for instance in speech disorders.

The realm of speech and language therapy involves several exciting ways in which researchers are trying to explore the full potential of ultrasound. Firstly, ultrasound can be a very useful diagnostic tool. Some speech disorders involve sounds that are not part of the language for other speakers, and it's not always easy to work out how such sounds are made, even for a trained expert. Ultrasound allows the clinician to see the tongue, providing clearer and more direct evidence about the nature of the underlying issue.

Another application is ultrasound visual feedback, where a patient undergoing speech therapy can see their own tongue movements while they are speaking. One such example is the production of the /t/ in 'tea', which involves raising the tongue tip towards the roof of the mouth, versus the production of /k/ in 'key', which involves raising the tongue back towards the rear of the mouth. Some children struggle to distinguish between these sounds, often producing the tongue tip and tongue back movements simultaneously. In order to investigate whether ultrasound visual feedback can help address such problems, Joanne Cleland of Strathclyde University and her colleagues tested using ultrasound feedback with children who had persistent speech sound problems. They provide evidence that ultrasound visual feedback can help some children to improve their pronunciation accuracy with these sounds. This is an active area of research in many countries, with researchers and speech therapists collaborating to find out whether ultrasound can be an effective tool in helping people with speech problems.

Future directions

Ultrasound has revolutionised our understanding of speech by providing an accessible and relatively low-cost tool for imaging the movements of the tongue. It already provides a huge potential for building new types of databases documenting different languages and accents. The increasing portability of ultrasound technology is also massively expanding its potential for studying less accessible communities. However, there is a lot of work to be done collecting, processing, and storing such data before they can be shared with the wider research community and the general public.

There are still many open questions and exciting avenues to explore. Ongoing advances in 3D ultrasound allow us to simultaneously obtain different views of the tongue, capturing its shape and movement in greater detail. In addition to this, researchers are constantly working on expanding the amount of useful information that can be extracted from the ultrasound images, giving us an even more powerful tool for analysing the dynamics of speech. ¶

Sam Kirkham is senior lecturer in phonetics at Lancaster University, where he founded the university's phonetics laboratory.

Patrycja Strycharczuk is senior lecturer in linguistics and quantitative methods at the University of Manchester. She has previously been awarded a British Academy Postdoctoral Fellowship.

Find out more

Web

Seeing Speech: a website with more information on ultrasound, plus example videos – seeingspeech.ac.uk

Dynamic Dialects: an interactive database of ultrasound videos collected from speakers of different English accents, including audio and ultrasound examples of the same words pronounced in different accents of English – dynamicdialects.ac.uk