

Why the human voice is so versatile

January 19 2017, by Noel Hanna



Credit: Andrea Piacquadio from Pexels

[Macaques](#) and [baboons](#) – two distantly related primates – are able to produce a similar range of voice-like sounds to humans.

In fact, many animals convey basic information using their [voice](#) but they don't display the full range of vocal abilities available to humans that enables our voice to be used for such a wide range of communication and entertainment.

This suggests that the uniqueness of the [human voice](#) is less in the anatomical ability to produce the sounds and more in our ability to precisely coordinate the physical movements, and to process the sounds into meaningful language.

Just how versatile is the human voice?

To get an idea of how versatile our voice is, we can think about how many intelligible sounds we make use of in a language.

Since English spelling is such a mess, this is more clearly illustrated by looking at the [pinyin romanisation](#) of Mandarin Chinese.

Using pinyin, Mandarin words can start with one of the following 24 sounds:

b, p, m, f, d, t, n, l, g, k, h, j, q, x, zh, ch, sh, r, z, c, s, w, y or nothing

These can be combined with the following 35 final sounds:

a, ai, an, ang, ao

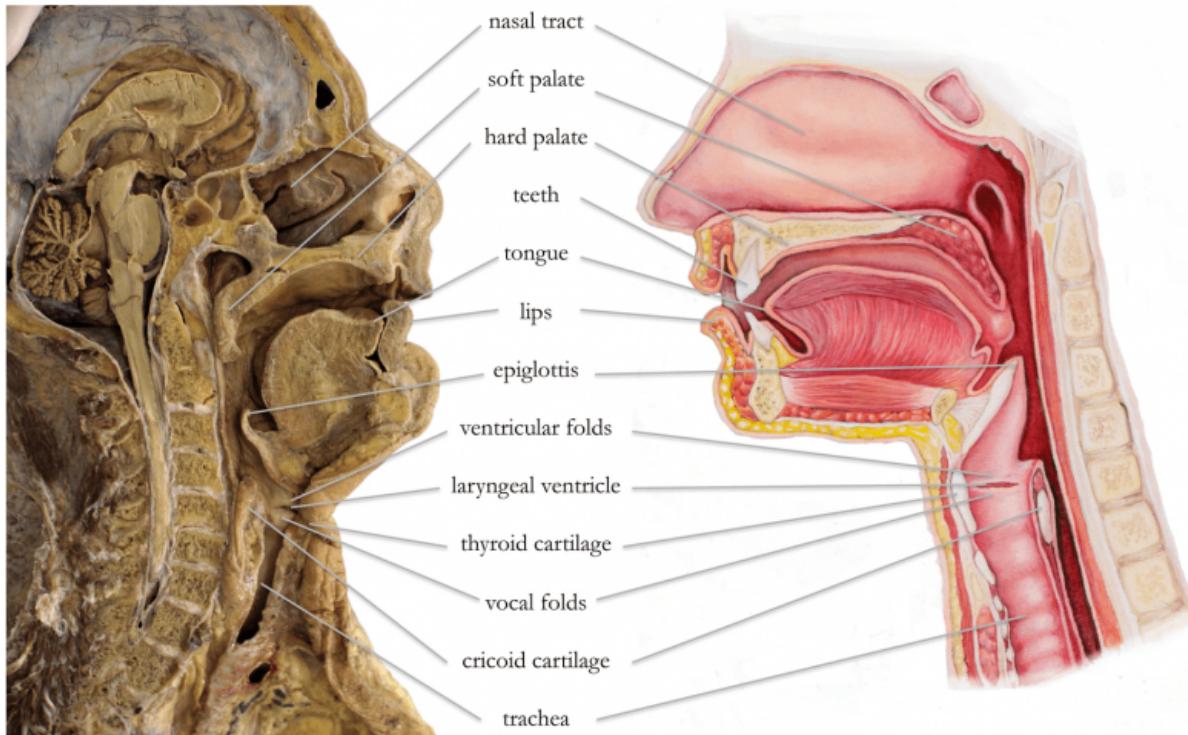
e, ei, en, eng, er

i, ia, iao, ian, iang, ie, in, ing, iong, iu

o, ong, ou

u, ua, uai, uan, uang, ui, un, uo

ü, üan, üe, ün



The anatomy of the voice. Credit: Noel Hanna, illustration Olivia Cox

This gives us $24 \times 35 = 840$ possible distinguishable sounds but each of these can have up to five tones (pitch patterns), which then gives us $840 \times 5 = 4,200$ unique words.

In practice, less than half of these words are actually used in the language.

But then you realise that most words in modern Mandarin are compounds of two of these words, so there are say $2,000 \times 2,000 = 4$ million possible unique words using this system of pronunciation, which are then strung together to make sentences.

And that is just one language. Each language has its own set of different sounds, which may or may not overlap with other languages.

So how do humans actually produce this variety of [speech sounds](#)?

How the voice is produced

Voice production can be thought of as a source-filter model. The voice is a combination of a vibrating source that controls its amplitude and pitch (the five tones in the example above), and an acoustic filter that controls how it sounds, much like how you can shape the sound with a graphic equaliser on a sound system.

The source is the vibrating vocal folds situated in the larynx. The filter is the airway that runs from the vocal folds to the lips or nostrils, which we call the vocal tract.

In the above image, the larynx (voice box) comprises the epiglottis to the cricoid cartilage. The thyroid cartilage tends to protrude from the neck in men and is called the Adam's apple.

Source: controlling pitch and amplitude

The vocal folds are two flaps of flesh that vibrate around 100-300 times per second (Hz) in speech.

The widely used name "[vocal cords](#)" came about from French anatomist

[Antoine Ferrein's analogy](#) that the air acted like a bow playing the strings (cordes in French) of the viola da gamba, or even a feather plucking the strings of a harpsicord.

While these analogies aren't very accurate, understanding the physics of vocal fold motion is still an active area of research, since experiments are so difficult. Observing the vocal folds is possible but not always practical. We can look at them but only from above – and even that isn't very comfortable.

In this example (video, above) the camera frame rate does not allow us to see the vibration of the vocal folds, but high speed video that shows the vibration is possible (video, below).

The vocal fold vibration isn't an on-off twitching of muscles, instead it is caused by the air that is passed over the vocal folds from the lungs. The frequency of vibration and its amplitude are controlled by a combination of pressure supplied by the lungs, the shape of the gap between the folds (the glottis), and the tension supplied by muscles in the larynx.

Learning to use all of these voice controls doesn't come easily – ask any teenage boy. Even singers take years to master the independent control of pitch and volume, which is put to the test by a practice technique called [messa di voce](#).

Filter: controlling articulation

Speech sounds, such as vowels and consonants, are determined by the vocal tract, which changes shape by moving the articulators (tongue, lips, soft palate, etc.) to filter the sound produced by the vocal folds.

Magnetic resonance imaging can give us a more detailed picture of the range of movement happening in the vocal tract (video, above) but it is

difficult to get three-dimensional information and it doesn't help us to see what the vocal folds are doing.

Although it is obviously more complicated, for a physicist, the vocal tract is something like a cylinder. It is a resonant system that is closed (or almost closed) at the vocal folds and open at the mouth.

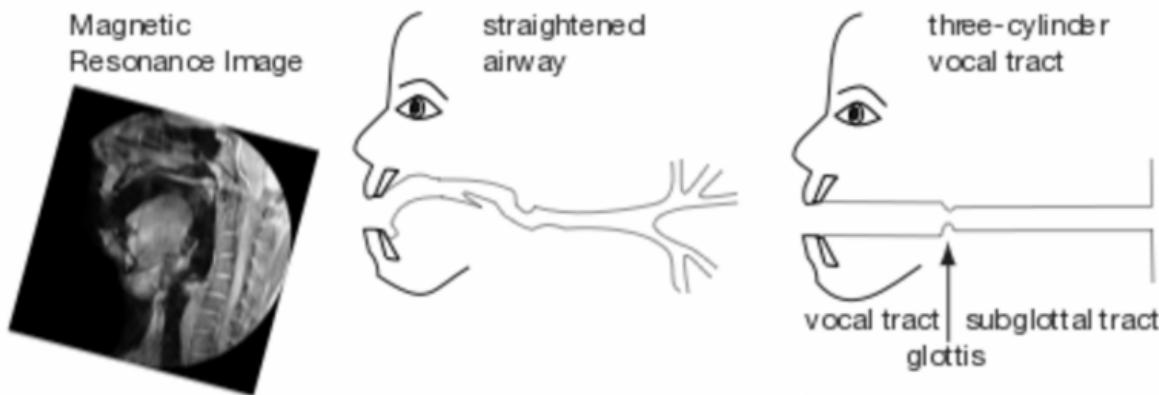
A resonant system allows standing waves to form. In the vocal tract the standing waves, or resonances, occur when the pressure is high at the vocal folds and low at the mouth.

The sound produced by the [vocal folds](#) at frequencies close to these resonances will be more noticeable. These more noticeable frequencies are called formants and they distinguish different vowel sounds.

For a 17cm long cylinder (about the length of a man's vocal tract) the first two acoustic resonances occur around 500Hz and 1,500Hz, close to what you would recognise as the vowel in the word "heard".

If you change the shape from a cylinder to [more realistic geometries](#) (by mimicking the effect of moving the articulators) then you change the position of the resonances, and therefore the vowel.

It may seem like a gross simplification to think of the vocal tract as a cylinder but in terms of acoustics this simple model allows us to determine the energy loss in the vocal tract.



Simplifying the respiratory airway helps us to understand its resonant properties.

Credit: Noel Hanna

It also gives us information about how rigid the walls of the tract are, as shown in [a paper by me and colleagues](#), which is important for producing plosive sounds like "p" and "b".

Learning

So if all humans (and some primates) can produce such a wide range of sounds, why do we have accents when we learn foreign languages?

Surely, if I want to learn Mandarin, I just need to train myself to produce those 2,000 sounds mentioned earlier. It would be almost like a form of physical exercise. The problem is our brains tend to categorise similar sounds. This hinders us in producing and perceiving sounds that do not fit into these categories.

For example, the French words for "above" and "below" ("dessus" and "dessous") tend to sound the same to untrained English speakers. When

we learn French, our brain must be taught to separate "u" and "ou" into two new categories, where previously there was only one.

So if our brains can't distinguish finely enough between the different sounds, could we use our understanding of [voice production](#) to improve language learning? Seeing the articulators inside our [vocal tract](#) in action is one idea that could help.

This example video shows a system that provides visual feedback on tongue and mouth movements to train pronunciation. Perhaps understanding voice production could help us to make those tricky new sounds when learning a language, and to further increase the versatility of our voice.

This article was originally published on [The Conversation](#). Read the [original article](#).

Provided by The Conversation

Citation: Why the human voice is so versatile (2017, January 19) retrieved 20 September 2024 from <https://phys.org/news/2017-01-human-voice-versatile.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.