# LING 190 Module 5: Origins of Speech

## Table of Contents



In this module, we'll be learning a bit about what *allows* humans to produce speech, despite how other closely related animals, like primates, cannot.

### 0.1. Why Can't Primates Speak?

Primates are the closest living relatives of humans. Early on, human scientists attempted to teach a chimpanzee how to speak, with very little success. The animal was able to learn 7 words in 6 years, but it was unable to match the speech ability of even a very young human child.

## 1. Animals make sounds!

### 1.1. Lecture Topics

- Many animals vocalize, in a variety of ways
- How have humans adapted our vocal tracts for speech?
  - Anatomy of the the vocal tract
  - Neural control over this anatomy
- How human anatomy and neural control, together, allow us to produce speech.

### 1.2. Birds Make Sounds!

Birds also have a trachea, like humans do, but bird tracheae typically split into two parts. Instead of a larynx, we say that birds have a syrinx. Some birds produce sound only on one side of this mechanism, while other birds produce sounds from both sides.

Some birds have whats known as a vocal organ, which allows them to produce some kind of sound by perturbing the airflow on either one or both sides of the syrinx.

## 1.3. Toothed Whales (Dolphins) Also Make Sounds!

Another common way of vocalizing, utilized by toothed whales (including orcas and dolphins) is to use their blowholes. They push air through their blowholes and use a structure known as the **anterior and posterior bursa** to achieve the same effect as how humans use the vocal folds.

Toothed whales, like dolphins, also have other parts of their head that change the characteristics of their vocalizations. The "Melon" of a dolphin can be tensed in certain ways to modulate the sound. Toothed whales can produce sounds at human-audible frequencies for social communications. In addition to these human-audible sounds, many toothed whales can produce sounds at very high frequencies for echolocation.

## 1.4. Apes (and many other mammals) Make Sounds!

We are particularly interested in how apes, and other mammales, make sounds, as opposed to humans.

- Just like humans, apes have a larynx, and basically all the other passive and active articulators that humans have, and can vocalize, but cannot speak.
- What is it that prevents them from being able to speak?
  - It's not just understanding concepts or the social significants of different types of communicative signs.

### 1.4.1. Historical Attempts to Teach Apes to Speak

- Early attempts to teach chimpanzees to speak (1930s to 1950s)
  - **Gua** (didn't learn to talk)
  - **Viki** (only learned 4 words)
- Early attempts to teach apes ASL (1960s to 1980s)
  - **Washoe** – A chimp who produced and recognized hundreds of signs and sign combinations
  - **Koko** – A gorilla who produced many, many signs and some sign combinations
  - **Nim Chimpsky** – Named after Noam Chomsky, a chimp who produced and recognized over 100 signs and some sign combinations

We were never able to teach apes to speak, but we were able to teach them to communicate with humans.

# 2. How have humans adapted our vocal tracts for speech?

There are two theories on why apes cannot speak:
1. The Anatomical Theory
2. The Neuro-behavioral Theory

## 2.1. When did we start to speak? (*phylogeny*)

**Phylogeny** – The study of evolutionary time.

We want to look at the category of primates, which includes humans, great apes (orangutans, chimpanzees, bonobos, gorillas), and in addition other species like lemurs, and lorises.

The anatomy of other hominins is very unclear, as we have no record of soft tissues in the vocal tract. However, we do know that some of these other species have similar bone structures in the area surrounding the vocal tract.

## 2.2. Anatomy of the the vocal tract

One idea is that what separates hominins from other hominids, is something about the anatomy that allows hominins like humans to produce speech but prevents other hominids like orangutans, or chimpanzees, from doing so. There are many of the same anatomical structures, but the shapes and sizes of these structures is all different.

1. One detail is that humans have a larynx that is very low, which makes the superlaryngeal vocal tract sound "larger."
    - Having a larger superlaryngeal vocal tract makes animals have a very low (deep) voice.
2. Humans can also control the tongue in 2 dimensions, whereas apes can only go front and back. Humans also have a large SLVT and can use it to produce more sounds.

### 2.2.1. But is this really a problem?

One way we can ask whether this anatomy really is a problem is to model the vocal tract of a monkey using an x-ray machine. We can use this model to simulate what would happen if the monkey could move its tongue in the same way a human can.

By doing this kind of study, many researchers have suggested that anatomy is not the problem. There is nothing, in theory, about the monkey vocal tract that prevents it from producing speech-like sounds (if the monkey had complete control over its tongue).

Conclusion: Anatomy is not the problem.

## 2.3. Sophisticated Neuromuscular Control over the Tongue and Other Superlaryngeal Articulators

The second theory of why humans can speak, but other great apes cannot, is that our brains have evolved to give us more sophisticated control over the muscles in our vocal tract, so that we can imitate vocalizations for social reasons. This change has specifically come about to allow us to produce speech.

Fitch (2000) makes some hypotheses:

1. Humans like to vocally imitate each other (just like parrots, seals, dolphins, etc.)
2. Being good at imitating the way someone speaks plays a social role (ability to identify in-group and out-group members)
3. Later became relevant for social communications

# 3. How does human anatomy and neural control, together, allow us to produce speech?

1. How can we synthesize speech just by knowing the size and shape of a vocal tract?
2. Introduce some new acoustic concepts (**harmonics**, **resonance**)
3. Source-filter theory and **formants**

## 3.1. How do we synthesize speech from a vocal tract?

First, we have to look at

1. Standing waves
2. Harmonics
3. Resonance
4. Source-filter theory

### 3.1.1. Waves and Harmonics

- Speech is often periodic, but it is never a “simple” waveform.
- How do we think about more complex waveforms?
- We’ll start with think about “sinusoidal” or “standing” waveforms.

#### 3.1.1.1. Standing Waves

- When waves hit a boundary, they can be reflected back.
- With the movement of different waves, you can have a stable parttern (standing wave).

#### 3.1.1.2. Complex Waves

#### 3.1.1.3. Complex Waves

- It’s possible to “add together” multiple standing waves of certain frequencies to create a complex waveform.
- Adding these waves together produces a stable (periodic) complex wave
  - Also known as a **composite sinusoidal wave**
- This is basically what most kinds of human speech (e.g. vowels) look like

#### 3.1.1.4. Harmonics

The standing waves that comprise our composite waveform are called **harmonics**. In speech, each harmonic’s frequency is an integer multiple of the lowest harmonic — known as the fundamental frequency — which corresponds to the rate at which your vocal folds vibrate. For example, if the fundamental frequency is 200 Hz, the second harmonic is 400 Hz, the third is 600 Hz, and so on. This fundamental frequency, often referred to as pitch, determines whether a voice sounds high or low.

We can visualize multiple harmonics on a two-dimensional graph, where the x-axis represents frequency in hertz and the y-axis represents the amplitude of each harmonic component. Each harmonic on this graph would be equally spaced, since they are all integer multiples of the fundamental frequency. Someone with a lower pitched voice would have harmonics that are closer together, while someone with a higher pitched voice would have harmonics that are further apart. It so happens that the lowest harmonics tend to be the loudest, and the highest harmonics tend to be softer.

### 3.1.2. Resonance and Source-Filter Theory

#### 3.1.2.1. Resonance

Resonance is a property of a material. If a wave comes into contact with a certain material or volume, and this material or volume has a certain **resonant frequency** that matches this wave, it will resonate and the wave will be amplified.

#### 3.1.2.2. Source-Filter Theory

The vocal tract can be thought of as **two** different air chambers that, ultimately, have different resonant properties (depending on how the tongue is shaped). What happens when we have control over these two air chambers is that we can amplify different harmonics that allow you to produce different vowel sounds.

Vowels are primarily characterized by the first few harmonics which are amplified by the two air chambers created by the tongue in the superlaryngeal vocal tract.

Acoustically, what is happening is that the larynx is producing a source, with a fundamental frequency and many harmonics, and the superlaryngeal vocal tract is filtering this source, and choosing different harmonics to resonate. What is produced is a complex waveform with different properties, and we can choose which properties are desirable, etc. Note that sometimes the harmonics, ordered from lowest to highest, may not have the same amplitude. It is even possible that $H_1$ will have a smaller amplitude than $H_2$. However, there is always an **overall** decrease in harmonic amplitudes.

What is to note, is that the specific harmonic that is amplified is not important for producing the desired vowel. What is important is the *frequency* of that harmonic. Therefore, regardless of the $F_0$ fundamental frequency of the speaker's voice, the filter used is the same, and the output waveform will look similar despite the density and location of the harmonics being different. This introduces a new concept, which is the **formant**.

#### 3.1.2.3. Formants

I.e. the frequenc(ies) of the harmonics that are amplified.

Formants are the resonant frequencies of the vocal tract. The first three formants, $F_1, F_2, F_3$, are the most important formants for speech sounds. For different vowels, the values of these formants will be different.