Morning everyone, welcome to our project presentation on the topic of **speaker recognition**.

This is Juliya, Shubham and I’m Shadow.

As the back Friday just passed, have you paid attention to the promotions **amazon** made on their **smart speakers**, or “Alexa”? This one got a price for lower than half during the **black Friday** last week. You can use this “Echo” to set timer, play Spotify, add calendar reminders or even make phone calls. But if there is **more than one user** at home, e.g. you and your mom. The speaker can **identify your voice** and add the reminder to **your calendar** instead of your mom’s. Your mom will not have a notification of doing a presentation today. How did it do that?

Or to say, **how did you do that**? You can do speaker recognition, right? How?

----------------------------------------------------------------------

Let’s look into **the ear**, your ear.

This is the ear structure in the textbook of **Gray’s Anatomy**. Yes, the TV series is called Grey’s anatomy after this textbook. You see sound is **collected** by the **ear**, then **guided** through this **ear canal**, then **passed** through this **membrane**. **Only vibration** can go through, **no direct flow** of air can pass. So this membrane works as a **capacitor** of **AC coupling**. Then the vibration was **transmitted** via these three **little bones**. They are **filters**. Then finally enter this **snail shell** structure, called **cochlea**. You can see these **yellow wires** goes into the snail shell? They are **nerves**, because the cochlea is where the **sound** is **converted** into **neural signals**. Let’s look into that.

------------------------------------------------------------------------

If we **pull** the snail shell **straight**, we can see a membrane in the middle of it. At the beginning, the membrane is **thin** and **stiff**, making the **resonant** vibration **frequency** very **high**. At the end, the membrane was **soft** and **lose**, making the **resonant** frequency very **low**. So, when a **sound travels** through this membrane, the **high** frequency components make **this** part vibrate, and **low** frequency components make **this** part vibrate. So what we have here is a **mechanical** version of **Fourier Transform**. These vibrations were **sensed** and **converted** to **neural signals** then send to the brain. And you are using this **Fourier transform** device **24 hours a day**.

-------------------------------------------------------------------------

So we **recorded** our voice, did a **Fourier Transform**, it looks like this. We can see the peaks indicating our **fundamental** and **harmonics** of our voice. So by looking at the FFT, we can **visually identify** the speaker, do the classification.

We **stacked** these three plots together and used as the **background** of this PPT.

We also fed this FFT result into **MATLAB neural network classifiers**. It gave us a **95%** accuracy. So up to this point, we can do the **classification by eyes**, and **MATLAB** can also do that, we believe our **data is good**. So we **moved on** to **build** our own **classifier**. **Shubham** will talk about how we built our classifier.