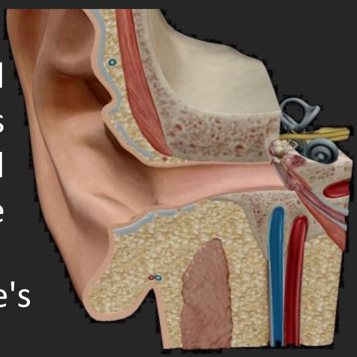
# Converting recorded voice to your OWN voice

**Team DU7**: Riadh Khlif; Yucheng Ni; Nikhail Virik; Ziyao Xiong **Supervisor:** Hugh Sparks

## 1. Background

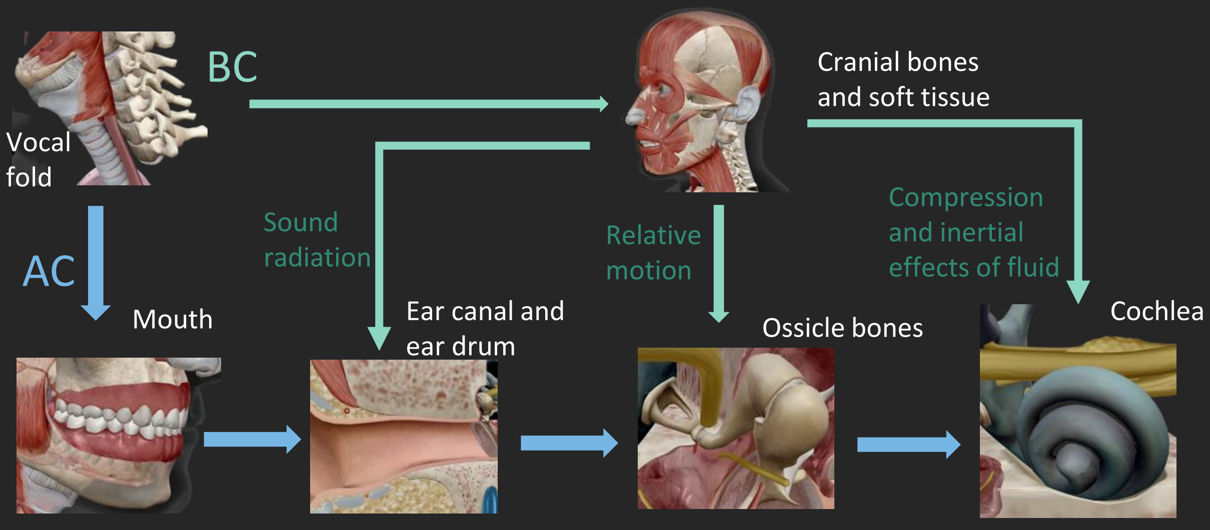
Surprised by how your voice sounds in recording? That's "voice confrontation," a disconnect between your expected and actual voice. Our self-heard voice blends air with bone-conducted sounds, enhancing lower frequencies and attenuating higher ones. For years, scientists have tried to replicate self-heard voices, mostly using subjective auditory perception or bone conduction microphones. Both ways have limitations: human auditory perception isn't exact, and the microphones can't reach all bone conduction pathways within the skull.



We plan to build a human head model in an acoustic physics field. By simulating bone and air conduction processes, we can create a reliable frequency response curve to replicate one's voice in recordings.

## 2. Pathway of bone conducted speech

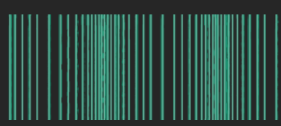
There are two primary mechanisms in our auditory system: air conduction (AC) and bone conduction (BC). AC refers to sound entering via the ear canal and being converted into neural signals after entering the cochlea. In contrast, BC occurs when sound, often via vibrations, reaches the inner ear through the skull. However, BC doesn't solely involve bone transmission: inner ear fluids and soft tissues are also involved, as shown in the figure below.




## 3. Theory and FEM Simulation

### *SOUND AS PRESSURE WAVES*

When sound propagates, the particles of the medium move in small oscillations around their equilibrium position, creating regions of higher pressure and lower pressure. Therefore, sound waves are essentially pressure waves, described by:



The wave equation: time domain

$$\nabla^2 P - \frac{1}{c^2}\frac{\partial^2 P}{\partial t^2} = 0$$

Fourier Transform

Helmholtz equation: frequency domain

$$\int (\bar{\nabla}\tilde{p}.\bar{\nabla}p).dV - \int\left(\frac{\tilde{p}p}{c^2}\right).dV = \int (j\rho_0\omega\tilde{p}q).dV - \int (j\rho_0\omega\tilde{p}\bar{v}.\bar{n}).d\Omega$$

Four terms in the Helmholtz equation can be described by its matrix form:

$$([K_a] + j\omega[C_a] - \omega^2[M_a]).\{p_i\} = \{F_{ai}\}$$

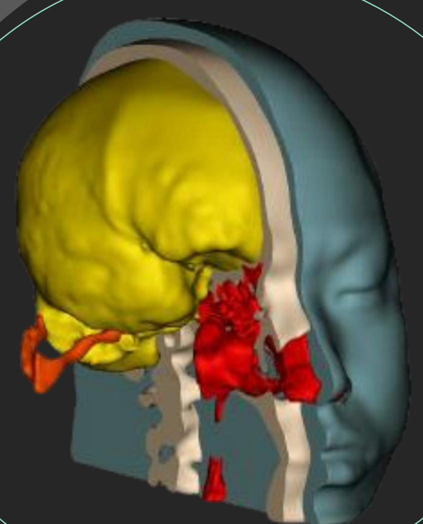
Stiffness to deformation of the acoustic medium

Damping or energy dissipation in the system

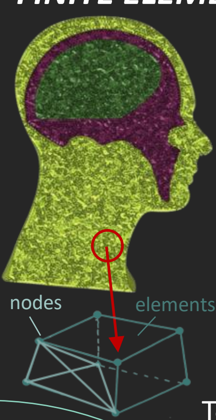
Inertia of the acoustic medium

External forces or sources that act on the system

Knowing the properties of the materials, we can calculate the sound pressure level at any point in a given sound field.



### *FINITE ELEMENT ANALYSIS*



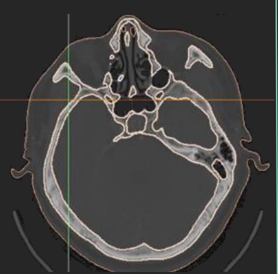

When a **DOMAIN** V is meshed, it is decomposed into a series of discrete **ELEMENTS** $V_e$. The elements in the mesh conform very well to the geometry. Adjacent elements are connected to each other at the **NODES**. The values of the field variable computed at the nodes ($p_i$) are used to approximate the values in the element interior (p) by interpolation of the nodal values that defined by a shape function Ni, such that

$$p(x,y,z) \approx \hat{p}(x,y,z) = \sum_{i=1} N_i(x,y,z).p_i \qquad (x,y,z) \in V_e$$

Together with calculation of Helmholtz equation at nodal points, the pressure in any element can be determined.

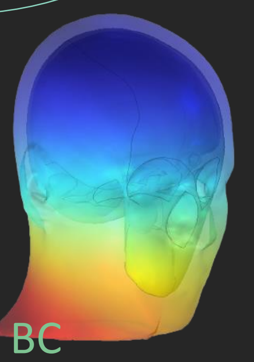
### *Running of Simulation*

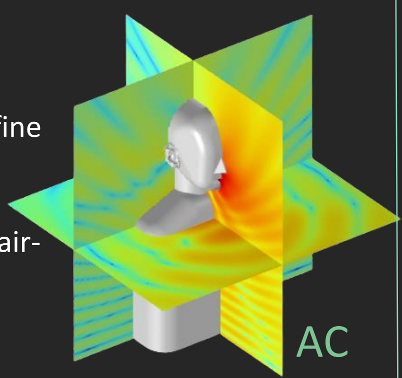
Step 1:
Using **Materialize Mimics** to convert the CT scan of a human head to a three-dimensional model.



Step 2:
Using **Hypermesh** to repair the model and discretize the model to tetrahedral elements.

Step 3:
Using **COMSOL** to define material properties in the model and apply bone-conducted and air-conducted simulation separately.



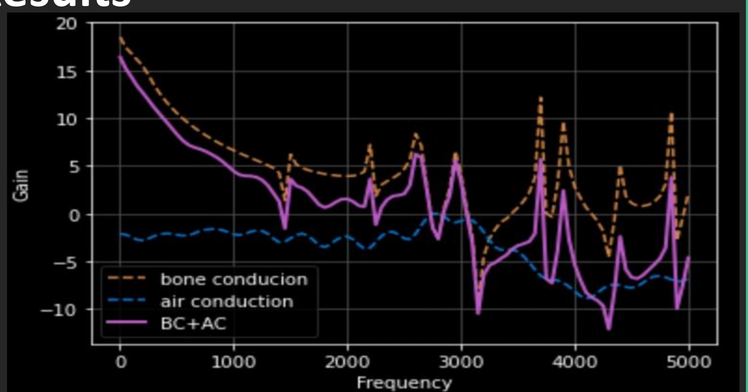

### *SOUND ATTENUATION IN MATERIALS*

The matrices depend on the factors listed in the table below. In general, sound loses energy through heat dissipation. This energy usually increases as frequencies increase, and less sound is absorbed in solids and liquids than in gases. In our simulation, the following materials were used (the soft tissue includes skin, muscles, and cartilages):

| Component | Young's modulus E (MPa) | Density ρ (kg/m³) | Poisson's ratio v | Loss factor η |
|---|---|---|---|---|
| **Brain** | 0.035 | 1000 | 0.45 | $3\times10^{-4}\times f$ |
| **Cortical bone** | 4000 | 2200 | 0.3 | 0.1 |
| **Soft tissue** | 0.7 | 900 | 0.45 | $3\times10^{-5}\times f$ |
| **Air** | / | 1.2 | / | / |

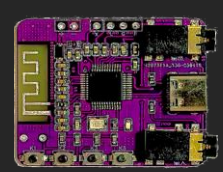
## 4. Simulation Results

As human perceive both air and bone conducted speech, we mixed each of their frequency response at a ratio 1:1. This assumption is based on a past research directed by Stefan Stenfelt: "Acoustic and physiologic aspects of bone conduction hearing".




## 5. Construction of DSP

A Digital Signal Processor (DSP) is a device that empowers us to incorporate the obtained transfer function and convert recorded voice directly into one's own voice through its input.



## 6. Conclusion

We managed to successfully obtain the transfer function of one's own voice. However, we weren't able to investigate the variations in BC effects among individuals, since we only utilized a single head model. We hope that the result we obtained serves as a generally representative model.

Imperial College London