Converting One’s Recorded Voices to One’s Own Voices by Acoustic Simulation

Project Team Code DU7: Riadh Khlif; Kevin (Yucheng) Ni; Nikhail Virik; Ziyao Xiong
Supervisor: Hugh Sparks

16 June 2023

Abstract

In this project, our objective was to simulate an individual's own voice. We achieved this by amalgamating the frequency response of air-conducted and bone-conducted speech, utilizing a CT-scanned model of a human head in a sound field. We assigned different materials to the various tissues within the head model and applied the finite element method to study the propagation behavior of acoustic waves. Furthermore, we incorporated the derived transfer function into a Digital Signal Processor (DSP), which allowed us to convert recorded voices into the individual's own perceived voice directly.

I. Background

The psychological phenomenon of "voice confrontation" reflects the disquietude people feel when their recorded voice doesn't align with their internal perception [1]. This discrepancy arises because the voice we hear combines air and bone-conducted sounds, enhancing lower frequencies and dampening higher ones [2].

Our project stemmed from a substantial mismatch between a speaker's self-perception and the recorded voice noticed during an English oral test. This discrepancy has considerable implications, as accurately reproducing one's voice is crucial in fields like language learning [3]. We therefore resolved to embark on a project on replicating one's self-perceived voices. However, current methods, whether through subjective auditory perception or bone conduction microphones, fall short due to their limitations in accuracy and accessibility [4]-[7]. To overcome these, we propose to construct a human head model and position it within an acoustic physics field. By simulating bone and air conduction processes, we aim to develop a reliable frequency response compensation curve, enabling more accurate voice reproduction in recordings.

We discussed the main points for which our project would be based with our supervisor, who suggested different pathways we could explore on the topic. We set our goals of the project according to his suggestions.

The initial goals consisted of:

- Thoroughly investigating the impact of bone attenuation and the impacts related to bone conduction.
- Implementing models of bone structures in a simulation software, bettering our understanding of body conducted speech to include it in our simulations.
- Obtaining the frequency response curve of both bone-conducted speech and air conducted speech, which can be combined to simulate one’s own voices.

Our goals extended to:

- Personalizing the algorithm to each of our group member’s voice.
- Integrating the frequency response into the Digital Signal Processor (DSP), where it can convert the voice directly from a recording.

II. Description of Project Work

Following our preliminary meeting with the supervisor on May 12th, we've established a routine of weekly meetings to review our progress and outline next steps. We have also arranged supplementary meetups between the four team members after lectures to collaboratively work on the computational components of our project. We've used OneNote lab book to document every meeting's content and the subsequent actions taken. Additionally, each team member has dedicated roughly 20 hours per week to independent work on the project.

A. Decision making

Our project leveraged the fact that bone conduction preserves more low-frequency sounds and attenuates high-frequency sounds [3]. Initially, we used the Equalizer APO software to manually adjust the frequency of our recorded voice. The success of this experiment affirmed the viability of our project.

Given the complexity of the human head and the challenges involved in manually calculating the sound field, we opted to apply modern technologies for analysis. Among two commonly used computational algorithms, the Finite-Difference Time-Domain (FDTD) and the Finite Element Method (FEM), we chose FEM. This decision was based on the fact that FEM is mesh-based, highly versatile, and aptly suited for handling complex geometries and frequency-domain problems [8]. There are several software options that employ the FEM, such as Ansys and COMSOL, each with its advantages and disadvantages. We decided to experiment with both. As for the model we intended to use, we explored the open-source database Visible Human Project [9], which provides CT and MRI scans of cadavers. However, these models lack open cavities and differ from live humans. Consequently, we chose to use a CT scan of a living person.

While we would have preferred to utilize MRI scans from multiple individuals as this would be more precise, budget constraints made this impossible. Thus, we decided to start with a CT scan of Ziyao's father's head, as he had recently undergone a medical examination.

B. Exploration of theory and initial data-taking

Ziyao established a Google Drive folder for the team to share important documents pertinent to our individual tasks in the first week. Utilizing these resources, Ziyao explored the bone-conduction pathways of sound, leading us to understand that bone conduction is influenced not only by the skull but also by different mechanisms through soft tissues [10]. This knowledge benefited Kevin and Nikhail in their role of applying finite-element analysis to analyze the acoustic behavior of human head. Similarly, Riadh's study on the mathematical framework underlying the linearized acoustic wave equation [11] was vital in the computational simulation of the frequency response. Ziyao and Nikhail further refined our understanding of the Finite Element Method, explaining to our supervisor how the meshing of shapes works in FEM.

Nikhail made progress by simulating a rudimentary head model using Ansys, assuming it as a sphere filled with water. Concurrently, Kevin researched the properties of various materials involved in a human head and secured a one-month free trial of COMSOL for further development.

C. Running of simulations

In the third and fourth week, Kevin built a 3D model of the head structures using the CT scan picture. This was done by the software 3D Mimic. In parallel, Nikhail examined and simulated the voice box using Ansys in order to identify the harmonic frequencies behind human vocalization. Integrating the simulated voice box in a rough head model allowed Nickhail to analyse the frequency response in different parts of the head, hence perfectly complementing Kevin's work on the interaction and propagation of sound in the surrounding structures. Using Riadh's notes on the fundamental and theoretical aspect of sound pressure and attenuation, Nickhail finally wrote a python code to simulate transmission and the attenuation of plane waves passing through layered media.

D. Pursuing extended goals and finalising results

The completion of the project during the final week was based on refining the simulation parameters by trial-and-error. This was manifested by the consideration of different meshing methods for our head model: we wanted to optimize the balance between high resolution and computational overhead. To obtain the coveted frequency response, Nikhail iteratively explored different shapes and adjusted the number of source points in the voice-box. To expedite our simulation process, Kevin procured four servers. This greatly contributed to us successfully deriving the frequency response curves, which accurately represented the behaviour of both bone-conducted and air-conducted speech. With these results, Kevin coded the transfer function into the DSP. As a team, we tested this to verify it operated as intended.

III. Summary of Results

Our main achievement is a frequency response curve combining both bone and air conduction, aligning with earlier research on this topic. Traditional relies on subjective auditory perception, using filters to adjust the frequency response, a process potentially flawed by human inaccuracies [4]. In light of this limitation, we believe our results offer more reliability. Additionally, we've constructed a compact DSP unit for presentations, enabling audiences to convert their voices directly, offering an engaging demonstration.

IV. Conclusion

We managed to successfully achieve the primary goals set for our project, including the extended objective of developing a DSP. However, due to time and budget constraints, we weren't able to investigate the variations in bone conduction effects among different individuals, since we only utilized a single head model for the simulation. While our research did not delve deeply into the individual differences, we hope that the transfer function we obtained serves as a generally representative model. We also recognized, in hindsight, that the project's progression could have been significantly boosted had we procured more powerful computer servers from the onset. As our model became increasingly refined, each simulation required hours to run, which highlighted the need for enhanced computing power.

V. Purchases, Existing Items and Use of 3D-Printing

A. Equipment used and purchased

As our project primarily involves computational work, we extensively used various computer software programs. They include:

- COMSOL and Ansys, for FEM analysis, 1-month free trial.
- Equalizer APO, for manually adjusting frequency response, open source.
- 3D Mimic, for converting CT scans to 3D models, 1-month free trail.
- Hypermesh, for constructing simulation grid, 50 pounds.
- Acpworkbench, for defining functions in DSP, 70 pounds.

We purchased the following items:

- Mvsilicon BP1048 DSP, audio mixer and acoustic amplifier unit for it, in total 20 pounds.
- Rents of four servers, 100 pounds.

We also made use of an existing CT scan.

VI. BIBLIOGRAPHY

1. Holzman, P. S., & Rousey, C. (1966). The voice as a percept. Journal of Personality and Social Psychology, 4(1), 79–86.

2. Békésy, G. V. (1960). Experiments in Hearing (p. 127). New York: McGraw-Hill.

3. Nakayama. (1997). Voice timbre in autophonic production compared with that in extraphonic production. The Journal of The Acoustical Society of Japan (E), 18, 67-71.

4. Rousey, C., & Holzman, P. S. (1967). Recognition of one's own voice. Journal of Personality and Social Psychology, 6(4), 464–466.

5. Yoram, M., & Hirose, K. (1996). Language training system utilizing speech modification. In Proceedings of ICSLP, pp. 1449–1452.

6. Shimada, T., & Imura, M. (2019). Technical support for vocal imitation by synthesizing one's own speech. In IPSJ SIG Technical Report (Vol. MUS-122).

7. Maurer, D., & Landis, T. (1990). Role of Bone Conduction in the Self-Perception of Speech. Folia Phoniatr Logop, 42(5), 226-229.

8. Finite Element Analysis Methods. (2010). Finite Element Analysis Concepts (pp. 1–22). https://doi.org/10.1142/9789814313025_0001

9. National Institutes of Health. (n.d.). The National Library of Medicine's Visible Human Project. U.S. National Library of Medicine. Retrieved from https://www.nlm.nih.gov/research/visible/visiblehuman.html

10. Stenfelt, S., & Goode, R. L. (2005). Bone-conducted sound: physiological and clinical aspects. Otology & Neurotology, 26(6), 1245–1261.

11. Jacobsen, F., & Juhl, P. M. (2013). Fundamentals of general linear acoustics. Chichester, West Sussex, United Kingdom: John Wiley & Sons Inc.