Use of Artificial Intelligence in Room Acoustics Prediction Using a Photograph

Daniel Milne | BSc (Hons) Audio Engineering | Supervisor: Lee Davison

1 Abstract

Acoustical parameters must be measured accurately to guarantee important acoustical qualities. However, acoustics measurement encompasses intricate methods, necessary training, high equipment costs, and is time-consuming (Papadakis and Stavroulakis 2018). This project details the research, creation, analysis, and testing of an Artificial Intelligence (AI) approach to RT60 estimation. A Convolutional Neural Network (CNN) has been created and incorporated into a usable product, capable of inputting up to 12 images of a space and analysing these images, to estimate the RT60 of the space. The CNN created achieved a validation accuracy of 38.5% per single photograph. In testing within three classrooms, the average estimate made by the AI achieved an average accuracy of within 8.98% (0.069 s) in comparison to an Interrupted Noise Method measurement; this is 29.1% (0.193 s) more accurate than the average human estimation. Due to time constraints, the testing data is limited to three rooms, of which the characteristics highly resemble the training data. Testing on unseen spaces on a wider scale must be performed, and to improve the accuracy and generalisability of the AI, more training data is required.

2 Aim and Objectives

Aim: To investigate if it is possible to create an alternative method for determining RT60 using AI.

Objectives:
- Understand the concept, capabilities, and intricacies of AI.
- Understand RT60, its meaning, methods of measurement, and calculation.
- Understand the impact certain acoustical characteristics have on people within the classroom.
- Research and understand how to program an AI system.
- Create an AI product usable by anyone, enabling RT60 estimation of classrooms.
- Research how the system can be improved upon.
- Test the accuracy of the system against humans.
- Use project management skills, including time management, and risk assessment, to safely complete the study on time.

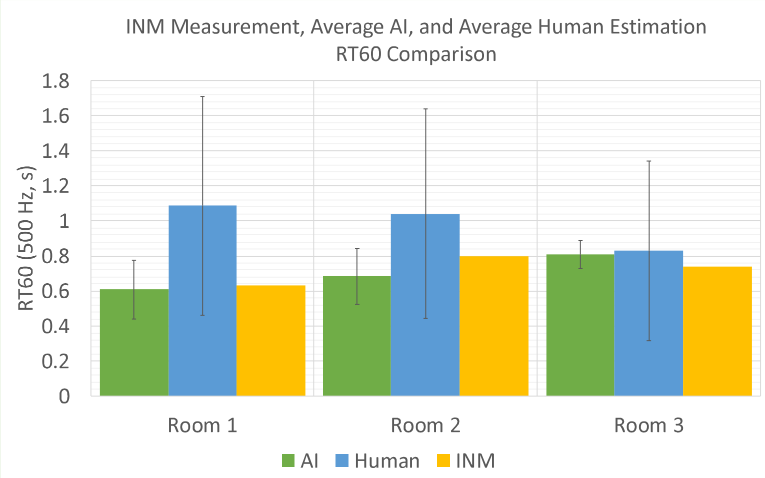

Figure 1 - Demonstration of the Higher Accuracy Achieved by the AI in RT60 Estimation Compared to Human Estimation

3 Method

- Potential risks, mitigation, and recommendations for establishing the Method, such as the use of pseudocode, were highlighted.
- The Balloon Impulse Response Method (BIRM), where a single balloon burst impulse response (IR) is recorded, was performed within 38 classrooms to ascertain RT60 of each room, and 24 photographs were taken within each classroom, totalling 912 photographs.
- The BIRM was validated for use as accurate data to train the AI, by testing the RT60 measured against that of the high-accuracy Interrupted Noise Method (INM) within three additional classrooms, at 500 Hz.
- A CNN was created, using parameters based on studies incorporating accurate CNNs. A software product was developed incorporating the CNN to estimate RT60, allowing a user to upload photographs of a space to the program, providing RT60 estimations.
- The product's estimation accuracy was tested within the aforementioned three additional classrooms, and compared against measured INM RT60 and human-visual estimation. The images the CNN sees were shown to participants of high-level experience with RT60, who were asked to estimate the RT60 of the spaces.

4 Results

- In testing, the BIRM achieved an average accuracy of within 15.67% compared to the INM RT60 measurements at 500 Hz.
- The CNN validation accuracy achieved was 38.5%.
- The product achieved an average accuracy of within 8.98% (0.069 s) compared to the INM measurements; this was 29.1% (0.193 s) more accurate than the average human estimation (Figure 1). There were 24 participants considered to have a high-level of experience with RT60.

5 Conclusion

For the first time, photograph analysis has been used to estimate room acoustics using AI and incorporated into a product (Figure 2). All objectives were met, and the system achieves higher accuracy than the average human estimation, however, to improve the product:

- More training data is required to improve the validation accuracy and generalisation of the CNN (Ciocca, Napoletano, and Schettini 2018, Gulli, Kapoor, and Pal 2019).
- Testing should be performed in rooms that are not with similar characteristics to the training data. Characteristics unrelated to the RT60 may affect estimations (Vanneman 2017), such as objects coincidentally being present in rooms of similar RT60.

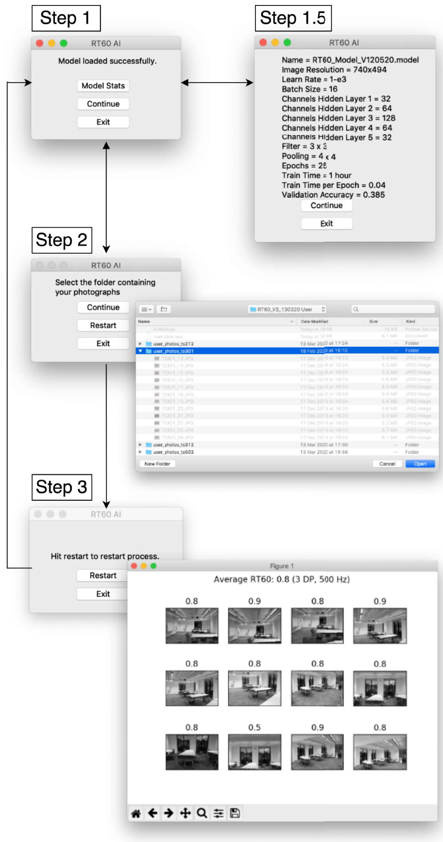

Figure 2 - Product User Interface

References

- CIOCCA, G., P. NAPOLETANO and R. SCHETTINI, 2018. CNN-based features for retrieval and classification of food images. *Computer Vision and Image Understanding (2018)*, 70–77
- GULLI, A., A. KAPOOR and S. PAL, 2019. *Deep Learning with TensorFlow 2 and Keras: Regression, ConvNets, GANs, RNNs, NLP, and more with TensorFlow 2 and the Keras API*. 2nd ed. Packt Publishing Ltd.
- PAPADAKIS, N. M. and G. E. STAVROULAKIS, 2018. Low Cost Omnidirectional Sound Source Utilizing a Common Directional Loudspeaker for Impulse Response Measurements. *Applied Sciences*, 8(9), 1703
- VANNEMAN, J., 2017. *Paper Club: Visualizing and Understanding Convolutional Networks* [09/05/20]. Available from: https://medium.com/@jamesvanneman/paper-club-visualizing-and-understanding-convolutional-networks-629fef9ceb8b