

USE OF ARTIFICIAL INTELLIGENCE IN ROOM ACOUSTICS PREDICTION USING A PHOTOGRAPH

D Milne Student Alumni, Solent University, Southampton, Hampshire, UK
L Davison Solent University, Southampton, Hampshire, UK
L Ausiello Solent University, Southampton, Hampshire, UK

1 INTRODUCTION

Current methods of determining acoustical parameters must adhere to current standards, require necessary training, include high equipment costs, and is time-consuming. The estimation of acoustics through calculation diminishes some issues with acoustics measurement but can be less accurate¹. It can be difficult for individuals and businesses in the acoustics field to consider RT60 despite its importance as an influential acoustic parameter² which arguably cannot be assumed, especially within educational spaces^{3,4}.

Artificial Intelligence (AI) is a program capable of performing tasks normally requiring human intelligence (see section 2.2) and has reached a maturity wherein it can now apply to real-world tasks that were not previously feasible within industries⁵. AIs might reduce costs as they can reduce the time needed to complete a task and reduce the number of resources required⁶. AIs require data, and in most cases, data acquisition, such as images⁷, video⁸, or text⁹, only needs to happen once which can then allow a range of applications including image identification⁷. As a proof of concept, this project aims to investigate if it is possible to create an alternative method for determining RT60 using AI.

2 BACKGROUND

2.1 RT60

2.1.1 What is RT60?

Reverberation can be defined as audible smearing in a sound signal created by indirect sound characterised by sound reflections within a space. RT60 is a reverberation time (RT) defined as the 'duration required for the space-averaged sound energy density in an enclosure to decrease by 60 dB after the source emission has stopped'¹⁰. RT60 can be measured or calculated at multiple frequencies to test if a space conforms to standards or recommendations.

There is no standard for the acoustics of University classrooms, however, standard BB93⁴, which applies to school classrooms, states that within furnished unoccupied classrooms intended for use by less than 50 people, the RT required is equal to or less than one second within refurbished rooms and equal to or less than 0.8 seconds within new builds, with the main aim of these standards to allow clear communication between lecturers and students, and students with other students⁴. Additionally, it was found that educational spaces with an RT60 of one second or higher resulted in the causation of 34% of students suffering anxiety within the classroom³.

2.1.2 Measuring RT60

The recommended methods in ISO 3382-1:2009 for measuring RT60 are the Integrated Impulse Response Method (IIRM) and Interrupted Noise Method (INM). The IIRM uses an omnidirectional sound source to create an impulse response (IR) that is not itself reverberant and has a sufficiently broad frequency range³. An IR is a temporal evolution of the sound pressure level observed at a position within a space as a result of the causation of an impulse at another position within the

space¹¹. Omnidirectional sound sources include the burst of a balloon or shooting of a starting pistol; however, omnidirectional standards, frequency-response standards, and repeatability cannot be guaranteed through such equipment¹². Dodecahedron speakers, which can also be used in the IIRM with a sine-sweep³, can be a highly accurate omnidirectional source instead, however, are typically expensive¹². The INM follows the same procedures as the IIRM with the exception that pink noise or white noise is instead played over an omnidirectional sound source, such as a dodecahedron speaker, and then stopped, and the time taken from the instance the noise is stopped until the point the sound pressure level drops by 60 dB is recorded as the RT60³. The aforementioned methods must be repeated, and measurements averaged to ascertain more reliable results. This is achieved by taking recordings with equipment in multiple positions to reduce the effects of room modes³.

2.1.3 Calculating RT60

Measuring RT60 requires resources and time, whereas calculating RT60 requires little resources yet is time-consuming; variables must be manually found for RT60 equation calculation, such as the volume and boundary surface area of a room, and the average absorption coefficients of the boundary surface materials¹³. There are multiple equations for calculating RT60, including the Sabine (Eq. 2.1), Norris-Eyring, and Mellington-Sette equations¹. These equations imply that RT60 increases as the size of the room increases and as the absorption of the room decreases¹⁴. The most accurate RT60 equation of those previously stated has been found to be the Sabine equation¹. Due to the minimal parameters involved, it can also be the most time-efficient equation for RT60 calculation.

$$RT_{60} = 0.161 \left(\frac{V}{\sum S\alpha} \right) \quad \text{Eq. 2.1}$$

Where:

V = Volume area (m³) S = Surface area (m²) α = Coefficient of absorption

2.2 AI

2.2.1 What is AI?

The first use of the broad term Artificial Intelligence (AI) originated in the 1950s¹⁵ and can be defined as meaning a program can perform tasks normally requiring human intelligence. AI is often broadly misunderstood to mean a computer system capable of completing a vast range of tasks a human can achieve, and everything a human brain can comprehend; this, however, is called Artificial General Intelligence (AGI). To date, there has been no complete implementation of AGI, and all current AI systems are referred to as Artificial Narrow Intelligence (ANI) as these systems have a limited scope of understanding, capable of only completing a specified task.¹⁶ Some ANI examples include image recognition⁷, speech recognition¹⁷, and classification¹⁸, however, AI has featured little in the audio and acoustics field.

2.2.2 Deep Learning

A major category of AI is Deep Learning (DL). DL systems use multiple layers of artificial neurons which store calculated values to predict an output. The development process of an AI program (Figure 2.1) involves analysing vast amounts of data to find meaningful structures within that data, allowing output prediction for unseen data¹⁹. For example, Ashqar and Abu-Naser's 2018 study⁷, where an AI program analyses images of tomato leaves for disease detection, includes a data set, called the Training Data¹⁹. This comprises images of leaves that are and are not diseased, called Features, plus the identification of a disease or no disease present on the image which is called Labels⁷. Once the Training Data is collected, AI systems 'learn' the relationship between the Features and the Labels through a mathematical process called 'training'. Then, unseen data can be presented to the AI and new predictions can be made. Accuracy of the system is often measured through Validation Accuracy;

the percentage accuracy achieved when testing the program against what is called testing data or reserved data, to see how accurately it can categorise unseen data it has not been trained for¹⁹. There is a significant issue with DL; much training data is required to be accurate, which can be time consuming to collect and difficult to source.

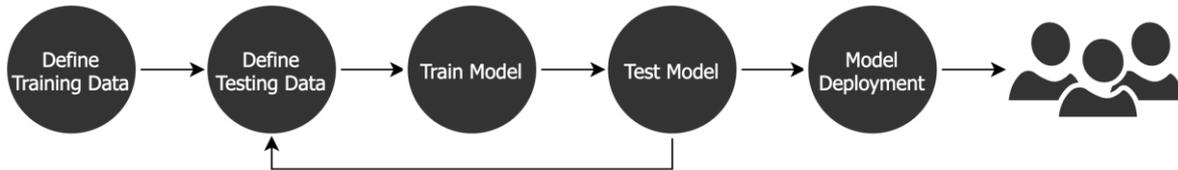


Figure 2.1 - The Development Process of a DL System (derived from¹⁶)

2.2.3 Convolutional Neural Networks

A DL system frequently used in image recognition is Convolutional Neural Networks (CNNs). CNNs are useful for analysing patterns in images to identify an object within the image or classify an image. Biological processes inspired CNNs; specifically, a paper by Hubel and Wiesel from 1962, where a significant advancement in the understanding of the operation of the visual cortex in mammals was achieved. This was the publication of a research paper that detailed the understanding of how light received by the visual cortex in cats activated different neurons within the brain, with specific neurons being more active in different areas depending on the position and orientation of a light source. It was detailed that the clearer a cat perceived an object, the greater the activity of neurons, resulting in the cat being able to recognise objects. This sign of neuron activity likely supported the later successful development of artificial equivalents.^{16,21,22}

3 METHOD

3.1 Data Acquisition

Data acquisition involves collecting the training data for the system, including photographs of rooms, for use as the Features, and an RT60 of that space, for use as the Labels. It is unknown how much training data is required for the CNN to be accurate. Within classrooms at Solent University, 24 photographs will be taken within each room to achieve an even symmetrical number of angles and, as a maximum of 38 rooms is available for this process, 912 photographs will be taken. An accurate measurement method, such as the Interrupted Noise Method would be ideally used, however, due to time constraints, a single balloon impulse response will be used to determine an RT60 within each of the 38 rooms, which will be referred to as the balloon impulse response method (BIRM). The use of the balloon bursts is to be justified/validated with an accuracy test through RT60 measurement of three additional classrooms using the BIRM and a highly accurate method for finding RT60; the INM¹⁰. Although information to calculate RT60 through the use of the Sabine equation may appear unnecessary due to the collection of RT60 data measured using the BIRM, it is justified as a data backup to mitigate any risks that may come from RT60 measurement.

After ensuring that the measurement microphone has been accurately calibrated using an NTi Class 2 calibrator, the method for data acquisition is as follows for each of the 38 classrooms, including the three additional classrooms to be used for BIRM validation:

1. Note the classroom name, the materials of each wall surface, and the measured room dimensions determined through using a laser measure.
2. Position the camera at the back of the room, facing one of the shortest sides.

3. Take a photograph level (0 °), slightly angled up, and slightly angled down (approximately 25 ° and -25 °).
4. Repeat step three moving equally clockwise around the room to take a total of 24 photographs.
5. Configure the Zoom H5 to a bit-depth of 24-bits, and a sampling rate of 48kHz, which are both sufficient to capture the desired frequencies accurately²³.
6. Attach the measurement microphone to an input of the Zoom H5 and configure the Zoom H5 to record from that input only and set the gain to a low-level, such as three.
7. Use hearing protection to prevent any balloons bursts from potentially damaging hearing.
8. Blow up a balloon so it is close to bursting, this is approximately 12 inches in height.
9. Slightly off-centre within the room and 1.2 metres above the ground¹⁰, position the microphone on a stand and press record.
10. Wait until any apparent noise has ended. At least approximately one metre away from the microphone¹⁰, burst a balloon near the centre of the room.
11. From the point the balloon was burst, wait five seconds, and then press Stop on the Zoom H5.

The Sabine equation is chosen to calculate RT60 due to the previously mentioned advantages. The coefficients used within these calculations are taken from Acoustic Project Company²⁴ due to the absorption coefficients provided being observably similar to the surface materials within the classrooms. The BIRM recordings are edited to remove unwanted information before and after the impulse response occurs, then normalised. After this processing, the impulse responses are imported into Room EQ Wizard, version V5.20 Beta 48, which is a program capable of IR analysis, to ascertain the RT60. The findings are then recorded into a spreadsheet at the frequencies provided by the program; 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, and 4 kHz.

The steps for data collection for the three additional rooms to be used for BIRM validation are performed using the INM. The same measurement microphone used in the BIRM is to be used but connected to the input of an NTi XL2 Acoustic Analyzer, and a Brüel & Kjær Dodecahedron speaker is used for an omnidirectional sound source, which has an inbuilt pink-noise generator. The following method of measurement adheres to industry standard ISO 3382-1:2009.

1. Ensure the signal output is muted and connect the signal generator to the speaker using an XLR cable and configure the generator to output pink noise.
2. Randomly position the speaker and microphone at least 1 m away from a wall, at least 2 m apart, and at the necessary distance from the floor; 1.5 m for the speaker and 1 m for the microphone.
3. Configure the XL2 for RT60 measurement and press the Play button.
4. Unmute the signal generator. After approximately five seconds and once the room has become 'excited', press mute.
5. Repeat steps 2 and 4 with the speaker and measurement microphone in new positions five times.
6. Press Stop on the XL2 and import the results to a computer.
7. Average the values of measurements for each frequency band by finding the mean value.

3.1.1 Programming

As this study is primarily aiming to develop a proof of concept and because of time constraints, the CNN will be trained to return a value at one frequency only; 500 Hz. This is the value frequently referred to when the RT60 of rooms is discussed due to this being a mid-frequency value²⁵. The parameters of the CNN model proposed are justified based on the successes of other published models^{17,26,27,28,29,30,31,32,33,34} and primary testing. Additionally, a minimum viable product (MVP) will be created, this is an interactive program allowing a user to input up to 12 images of a single room. It will then return an average RT60 value of that room to the user.

3.2 Testing

3.2.1 INM Measurement, AI Prediction, and Human Guess Comparison

The Validation Accuracy of the AI is calculated and recorded by the AI itself during the development phase. However, a real-world test is more important and interpretable, and so the accuracy of the final product will be tested in a scenario similar to how it will be used by an end-user. A comparison will be made between the AI's prediction, the INM measurement, and the guesses of humans with a high-level of experience in RT60. As the data is available, a comparison to the RT60 calculated through the Sabine equation will also be included. This test will be performed in the same three classrooms the BIRM validation test occurs, meaning the AI has not seen the test rooms before. Using the same camera as during the data acquisition phase and at similar angles, 12 photographs are to be taken and imported into the AI program. The average RT60 will be predicted by the program calculating the mean average RT60 from the 12 photographs.

3.2.2 Collecting Human Guesses

The RT60 estimation from humans with a high-level of experience with RT60 will be gathered through an online survey to compare the abilities of the AI and a human. This would mean showing participants the greyscale and low-resolution images that the AI is presented. Strict protocols must be followed to reduce the risk of ethics impact, this is achieved by making the results anonymous and allowing certain demographic questions to be unrequired to complete the survey. It is imperative that results only consider participants with a high-level understanding of RT60 to improve the validity of the human guesses. To achieve this, simple questions will be asked about RT60 and demographic data about the participant's experience in acoustics will be gathered to test if the participant is to be considered. Exact requirements for qualification cannot be defined at this stage of the project, as it is unknown how many participants will partake, and many participants are required to validate results³⁵.

4 RESULTS

4.1 BIRM Validation

Table 4.1 shows that within the test rooms, in comparison to the INM measurement for RT60 at 500 Hz the BIRM achieves an average accuracy of 84.33%, this is an accuracy within 0.113 seconds.

Table 4.1 - The Accuracy Difference Between RT60 Measurement Methods at 500 Hz

Room	INM RT60 (s)	BIRM RT60 (s)	Sabine RT60 (s)	INM to BIRM Difference (s)	INM to BIRM Accuracy (%)
1	0.630	0.533	1.214	-0.097	84.61
2	0.800	0.694	1.246	-0.106	86.75
3	0.740	0.876	1.246	+0.136	81.62
Mean Average				0.113	84.33

4.2 AI Accuracy

The final CNN model achieved a Validation Accuracy of 38.5%. Table 4.2 demonstrates how the AI predicted RT60 through averaging the individual estimations for the 12 images for each room.

Table 4.2 - Averages of the AI's RT60 Estimation at 500 Hz

	RT60 (s)		
	Room 1	Room 2	Room 3
Range	0.5	0.4	0.3
Mode	0.5	0.5	0.8
Mean	0.608	0.683	0.808

4.3 AI Versus Human

The survey was completed by 36 participants, of these, 24 were considered having a high-level understanding of RT60. The 12 participants who do not meet this standard are omitted from the results and justifications for removal are discussed in section 6.4. Within Table 4.4 and Figure 4.1, it is demonstrated that, in comparison to the INM RT60 measurement at 500 Hz, the AI achieved an accuracy of 90.90%, this is an accuracy within 0.069 seconds. This is 28.98% or 0.192 seconds more accurate than the average accuracy achieved by the human guesses. Figure 4.1 also shows that the AI has a significantly lower range of individual errors in comparison to that of human participants.

Table 4.3 - INM Measurement, AI Predicted Mean Average, and Human Guessed Mean Average for RT60 at 500 Hz

Room	INM Measurement (s)	AI Predicted Average (s)	Human Guessed Average (s)
1	0.630	0.608	1.085
2	0.800	0.683	1.041
3	0.740	0.808	0.828

Table 4.4 - INM to AI Prediction Comparison and INM to Human Guess Comparison for RT60 at 500 Hz

Room	INM to AI		INM to Human	
	Difference (s)	Accuracy (%)	Difference (s)	Accuracy (%)
1	-0.022	96.51	+0.455	27.78
2	-0.117	85.38	+0.241	69.86
3	+0.068	90.81	+0.088	88.11
Mean	0.069	90.90	0.261	61.92

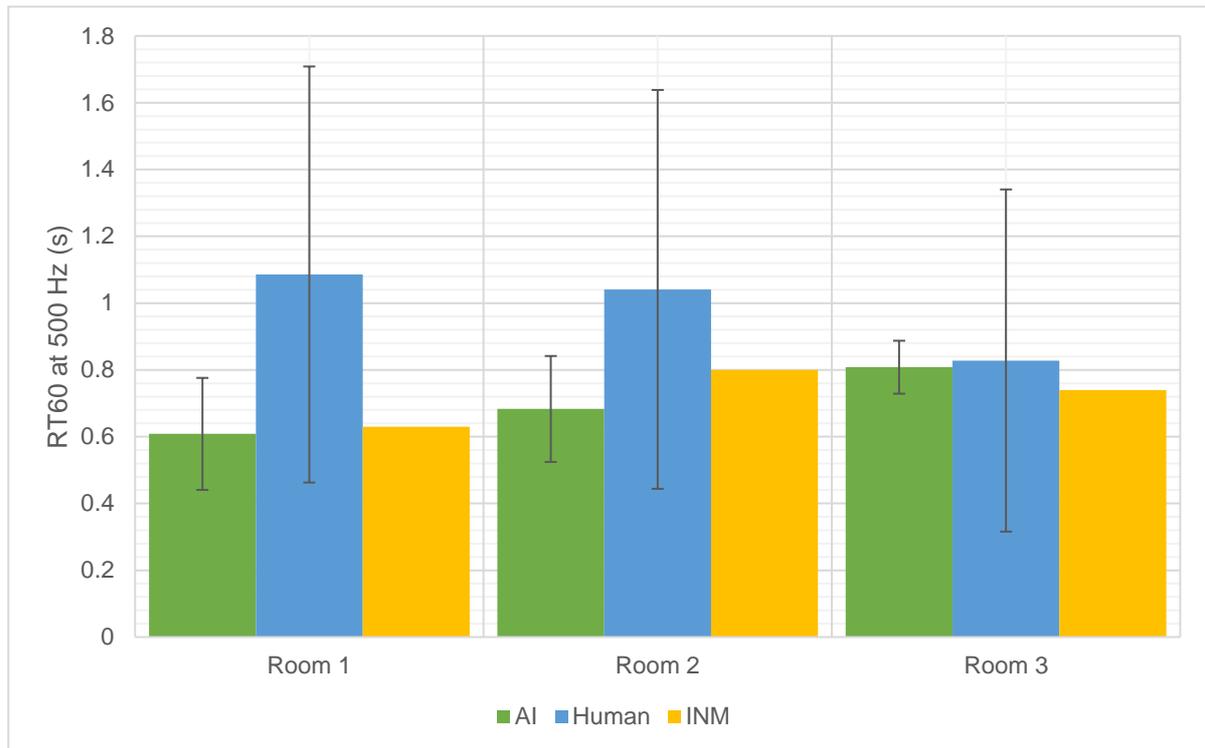


Figure 4.1 – Comparison of the Mean-averaged AI Estimation, Mean-averaged Human Guess, and INM Measurement

5 DISCUSSION

As discussed previously, 912 photographs of classrooms were taken with 38 associated IRs recorded within each room using the BIRM. To validate the use of the BIRM, a comparison to a high-accuracy method of RT60 measurement had to be made; using an additional three classrooms, BIRM measurements were taken following the same methods used in data collection and INM RT60 measurements were made. The results found that the BIRM achieved an average accuracy of 84.33% compared to the INM across the three rooms, this is approximately an accuracy within 0.113 seconds.

For the same three classrooms, the AI predictions were then compared to the INM measurements. It was found that the averaged AI prediction achieved an average accuracy of 90.90% compared to the INM measurement across the three rooms, this is an accuracy within 0.069 seconds. The AI estimations were then compared to human estimations of the same three classrooms, and it was found that the AI could estimate with an accuracy 28.98% or 0.192 seconds higher than the average human estimation after analysis of the same images. Figure 4.1 shows a significantly higher standard deviation for human estimations compared to AI estimations, demonstrating that the range of human estimations is much higher than that made by the AI. The validity of findings is discussed in the Evaluation section.

6 EVALUATION

6.1 Data Acquisition

Table 4.1 shows that a BIRM measurement for RT60 at 500 Hz is on average within 0.113 seconds of an INM measurement, this is an accuracy of 84.33%. Ideally, testing the accuracy of a balloon

impulse response would be performed in more than three rooms to increase the validity of results and unfortunately, due to the availability of classrooms at the time of testing, only rooms of RT60 between 0.63 and 0.8 seconds were available. As the AI is trained on rooms of RT60 between 0.4 and 1 seconds these findings do not represent the RT60 range the AI can predict, meaning that estimated RT60 results outside the tested region cannot be validated. There is not sufficient data to argue persuasively that a balloon impulse response will always be within 0.136 seconds of a highly accurate RT60 measurement at 500 Hz, as was found here. Additionally, when used as impulse responses, it is known that balloons cannot achieve sufficient dynamic range at low frequencies, and the larger the balloon the greater the amplitude is likely to be³⁶; therefore, balloon size was kept consistent throughout the experiment, but any human error or any distortion in the balloon's material could reduce the likelihood of achieving a consistently high-dynamic range, potentially reducing the accuracy of measurement in that instance. Furthermore, to achieve higher accuracy, measurements should be made multiple times and averaged when measuring RT60¹⁰, due to the single balloons impulse responses used to measure the RT60 of each room, there may be an irregular noise during the recording of the IR that was not noticed, distorting the IR measurement.

The results in Table 4.1 show that training the AI on RT60 values calculated through the Sabine equation would lead to erroneous prediction due to its high inaccuracy, which is sometimes close to double the INM measurement. However, these findings regarding the Sabine calculation are not conclusive as, due to time-constraints, individual features of rooms such as doors, televisions, and whiteboards were not included within the classroom absorption coefficients. Additionally, absorption coefficients used within Sabine calculations referenced the Acoustic Project Company²⁴, which included some differences to the actual surface material type. Furthermore, it is unknown how accurately the absorption coefficient was determined for the data provided, potentially leading to higher inaccuracy when using the Sabine calculation.

6.2 AI Accuracy

When averaged over 12 images, the accuracy of the AI's predictions is higher than that of a human guess for all three rooms (Table 4.4). However, the Validation Accuracy is low compared to other studies that use similar CNNs^{7,30}; likely due to the larger data sets employed by these studies. More training data leads to generalisability, and the AI could likely closely estimate the RT60 of the three test rooms because the testing and training data was not sufficiently broad, as similar rooms have been used to train the AI, meaning that there may be characteristics unrelated to RT60 that the AI has used to learn the RT60 of a space. For example, the AI could have been trained on rooms with similar RT60's that have features unrelated to RT60, such as the carpet pattern being the same. It could be tested if this is the case by altering aspects of the room that do not affect RT60, or digitally removing aspects of an image that should be unrelated to the output and seeing if the AI's estimated RT60 changes with each alteration³⁷. To improve the generalisability and accuracy of the AI, many photographs of rooms should still be used, but with additional photographs of the same classrooms taken with slight alterations to characteristics that should not influence RT60 such as, for example, the positioning of ceiling lights, writings on whiteboards, or wall paint blemishes and markings.

6.3 AI MVP Implications

As discussed in the previous section, it is hoped that with more training data the product will offer a service that can predict RT60 as accurately as when measured using the INM. Furthermore, aside from the primary aim of predicting the RT60 of classrooms, particularly at universities to ensure a satisfactory RT60 similar to that in standard BB93⁴, with high accuracy this product could be implemented within a range of scenarios such as estimating in-game RT for rooms; a programmer will no longer need to configure invisible zones within games to trigger different RT within different rooms and environments. Potentially a version of the product could be built into the game engine, estimate the RT within an in-game room and automatically make any sound effect that occurs within that room have the correct RT time applied similar to what would be experienced in real life. Professional music studios could be analysed also to easily find if a studio conforms to a

recommended RT. Photographs of rooms could be digitally edited and then input to the program to see how RT60 changes due to new furniture, room layout alterations, or added acoustic foam panelling before making a purchase.

6.4 Human Guesses

What is considered being a human with a high-level of experience with RT60 was determined by ensuring the participant had greater or equal to five years of experience within the audio and acoustics field, had measured RT60 at least once, and had qualifications of a higher education level or higher. Unfortunately, the number of participants is too low to confidently state that any erroneous guesses are not representative of the group. Additionally, one-third of participants stated that they have or may have been within these rooms before, potentially resulting in their estimations being different to those who have never been in the room; having audibly experienced sound characteristics is likely to improve the accuracy of the participants' guesses. Furthermore, the high standard deviation by humans compared to the AI shown in Figure 4.1 may be more pronounced because of the limited range the AI could estimate within and the unlimited range a human could estimate within. The AI makes estimations between 0.4 and 1.0 seconds, therefore, participants should have also been limited to the same range for a fairer comparison.

Photographs used as training data were taken within a UK university, however, only 37.5% of participants live or spend most of their time within the UK meaning that 62.5% of participants may have less experience within rooms of this design or architecture, potentially resulting in greater difficulty in guessing the RT60. It may have been more useful to ask participants if they have been in a UK university classroom before, as this would indicate if the participant had audibly experienced the characteristics of similar rooms. Furthermore, the survey language was not translated from English, reducing non-English speakers' chances of equal representation within the study³⁸.

6.5 Techno-cultural Implications

One source has claimed that although AI may globally displace 75 million jobs by 2022, 133 million new roles will be created³⁹, and another has stated that by 2037 the UK will have created 7.2 million new jobs, displacing 7 million jobs⁴⁰. Non-AI job automation throughout the past two centuries has had greater detriment to the low-skilled or semiskilled workforce, however, the transition to higher-skilled jobs has occurred in the long-term⁴¹. In the modern era, a transition to higher-skilled roles needs to be provided, like has been seen by Amazon, who has announced \$700 million will be used to train approximately 100,000 workers in the USA by 2025 to support a transition to higher-skilled jobs⁴².

Using AI in the workplace to increase automation use can increase productivity, resulting in reduced costs and increased sales for a business, which can create a demand for additional workers because of the business being more financially stable to invest in new areas⁴⁰. It is thought that employees within this industry will feel more valued by the company they work for, as more jobs are likely to require skills such as creativity and teamwork that machines find challenging to replicate⁴⁰. It is also hoped that job automation can lead to a reduced working week, something that has been argued to be beneficial for human happiness/wellbeing and the economy⁴³. It has been found that 55% of women and 47% of men state they 'always' or 'often' return from work exhausted⁴³, similarly, 70% of nurses have been found to be needing to work 'very hard' frequently⁴³. It is recommended that further research be performed to support the use of AI within industries, as it remains unclear how concerning the implications will be regarding wages, productivity, and jobs, as the future impact of AI and job automation varies between sources.

7 CONCLUSION

The original project aim was to investigate if it is possible to create an alternative method for determining RT60. For the first time, photograph analysis has been used to predict room acoustics. Although the AI MVP was developed as planned and was able to estimate RT60 with an average accuracy within 0.069 seconds or 90.90% and with greater accuracy than the average guess by humans with a high-level understanding of RT60 using the same resources, it cannot be confidently stated that the findings of this study validate the accuracy of the AI without further testing. A vital future step is the collection of more training data to increase the accuracy of the AI. With more training data, the potential range of AI RT60 outputs can also be increased, which is important to allow the system to identify classrooms with an RT60 higher than one second, which is deemed uncomfortably high³. Furthermore, testing of CNN model architecture needs to be improved to understand what the best architecture is for this system to achieve the highest accuracy. Due to the black box nature of CNNs²⁰, it is unknown how RT60 predictions were made by the AI, and more testing would need to be performed on the current system to conclusively state that the current system can predict RT60 accurately in a range of rooms. Despite these criticisms, it is hoped that this study has enabled further development in this field, proposed ideas for the future implementation of this concept, and demonstrated the issues of the system with accompanying future work required to make a final product that can be beneficial.

8 REFERENCES

1. PUTRA, J., RABIYANTI, and I. RAHMANIAR, 2018. Evaluation of Reverberation Time of Class Room. AIP Conference Proceedings 1977, 020042 (2018)
2. DUMORTIER, B. and E. VINCENT, 2014. Blind RT60 estimation robust across room sizes and source distances. 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)
3. ATHARI, P. *et al.*, 2014. The Importance of Acoustic Quality in Classroom. Jurnal Teknologi. 70:7, 71–76
4. EDUCATION FUNDING AGENCY, 2015. BB93: acoustic design of schools - performance standards
5. SZOLOVITS, P., 2019. Artificial Intelligence in Medicine. Routledge.
6. BERMEJO, P.H.D.S. *et al.*, 2019. How and where is artificial intelligence in the public sector going? A literature review and research agenda. Government Information Quarterly, 36(3), 385-624
7. ASHQAR, B. A. M. and S. S. ABU-NASER, 2018. Image-Based Tomato Leaves Diseases Detection Using Deep Learning. International Journal of Academic Engineering Research (IJAER), 2(12), 10-16
8. FRIDMAN, L. *et al.*, 2016. Driver Gaze Region Estimation Without Using Eye Movement. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019.
9. BUDIAWAN, R., P. TELNONI, and M. QANA, 2019. Comparison of Machine Learning Classification Method on Text-based Case in Twitter. 2019 International Conference on ICT for Smart Society (ICISS)
10. BS EN ISO 3382-1:2009. Acoustics — Measurement of room acoustic parameters
11. ISO 354: 2003. Acoustics—Measurement of Sound Absorption in a Reverberation Room; ISO: Geneva, Switzerland, 2003
12. 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
13. MIDDLETON M. W., 2002. Reference Data for Engineers. 9th ed. Elsevier. 40-1 to 40-28
14. ROSSING, T., 2007. Springer Handbook of Acoustics. Springer Science & Business Media, 393-394
15. KAPLAN, J., 2015. Artificial Intelligence: What Everyone Needs to Know. Oxford University Press, 18

16. WITMAYER, C., 2019. Automating Metadata Logging Through Artificial Intelligence. SMPTE Motion Imaging Journal, October 2019
17. ALHAWITI, K. M., 2015. Advances in Artificial Intelligence Using Speech Recognition. International Journal of Computer, Electrical, Automation, Control and Information Engineering, 9(6), 1439-1442
18. COSTA, M. G. *et al.*, 2016. Combining visual and acoustic features for bird species classification. 2016 IEEE 28th International Conference on Tools with Artificial Intelligence
19. 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.
20. BAG, R. *et al.*, 2019. Social Network Analytics Computational Research Methods and Techniques. Academic Press. 2019. 127-147
21. CORNELISSE, D., 2018. An Intuitive Guide to Convolutional Neural Networks [17/09/20]. Available from: <https://www.freecodecamp.org/news/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050/>
22. HUBEL, D and T. WIESEL, 1962. Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. The Physiological Society, Jan 1962.
23. BALLOU, G., 2015. Handbook for Sound Engineers. 5th edition. Burlington, MA: Focal Press. Page 13
24. ACOUSTIC PROJECT COMPANY, 2020. Reverberation time calculation [17/09/20]. Available from: http://www.akustik.ua/calculator_rt60_en.html
25. JONES, G, D. LAYER, and T. OSENKOWSKY. National Association of Broadcasters Engineering Handbook: NAB Engineering Handbook. Taylor & Francis. Page 405.
26. AHMED, F. *et al.*, 2016. Reducing Overfitting in Deep Networks by Decorrelating Representations. ICLR 2016.
27. FAVORSKAYA, M. and A. PAKHIRKA, 2019. Animal species recognition in the wildlife based on muzzle and shape features using joint CNN. Procedia Computer Science, 159, 933-942
28. ANTONOGLU, I. *et al.*, 2019. Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. ArXiv 2019.
29. CIREŞAN, D. *et al.*, 2011. Flexible, High Performance Convolutional Neural Networks for Image Classification. Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, 2011.
30. BERA, T. *et al.*, 2019. A deep learning model to recognize food contaminating beetle species based on elytra fragments. Computers and Electronics in Agriculture 166 (2019), 1-8
31. KIM, H. *et al.*, 2017. Comparison of Shallow and Deep Learning Methods on Classifying the Regional Pattern of Diffuse Lung Disease. J Digit Imaging (2018), 415–424
32. 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
33. KHANDELWAL, R., 2019. Overview of different Optimizers for neural networks [17/09/20]. Available from: <https://medium.com/datadriveninvestor/overview-of-different-optimizers-for-neural-networks-e0ed119440c3>
34. MASTERS, D. and C. LUSCHI, 2018. Revisiting Small Batch Training for Deep Neural Networks. Graphcore Research.
35. CHAPMAN, S. and P. MCNEILL, 2005. Research Methods. 3rd ed. Routledge.
36. DOMITROVIC, H., M. HORVAT and K. KAMBROSIC, 2008. Reverberation time measuring methods. Acoustics 08 Paris. Pages 8511 – 8516
37. VANNEMAN, J., 2017. Paper Club: Visualizing and Understanding Convolutional Networks [17/09/20]. Available from: <https://medium.com/@jamesvanneman/paper-club-visualizing-and-understanding-convolutional-networks-629fef9ceb8b>
38. CHOI, B. and A. PAK, 2005. A Catalog of Biases in Questionnaires. Prev Chronic Dis. 2005 Jan; 2(1): A13.
39. WORLD ECONOMIC FORUM, 2018. The Future of Jobs Report. Page 8

40. BBC, 2018. AI will create as many jobs as it displaces – report [17/09/20]. Available from: <https://www.bbc.co.uk/news/business-44849492>
41. AUTOR, D., 2015. Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*. Volume 29(3), Pages 3–30
42. PRESS, G., 2015. Is AI Going To Be A Jobs Killer? New Reports About The Future Of Work [17/09/20]. Available from: <https://www.forbes.com/sites/gilpress/2019/07/15/is-ai-going-to-be-a-jobs-killer-new-reports-about-the-future-of-work/#43646ca9afb2>
43. HARPER, A. and W. STRONGE, 2019. The Shorter Working Week: A Radical and Pragmatic Proposal. Page 29