

A PRACTICAL SYSTEM TO PREDICT THE ABSORPTION COEFFICIENT, DIMENSION AND REVERBERATION TIME OF ROOM USING GLCM, DVP AND NEURAL NETWORK

M.N. Yahya^{1,2}, T. Otsuru², R. Tomiku², T. Okuzono²

¹Faculty of Mechanical and Manufacturing Engineering, UTHM
86400, Parit Raja, Batu Pahat, Johor Malaysia
Phone: +607-453-7742; Fax: +607-453-6080
Email: musli@uthm.edu.my

²Dept. of Architecture and Mechatronics, Faculty of Engineering, Oita University,
700 Dannoharu, Oita 870-1192, Japan
Email: otsuru@oita-u.ac.jp, tomiku-reji@oita-u.ac.jp, okuzono@cc.oita-u.ac.jp

ABSTRACT

Various prediction techniques of reverberation time such as the Sabine and Eyring equations, ray-method, and numerical method require main parameters such as the absorption coefficient and dimensions. Normally, these parameters are obtained from references or/and measurements that necessitate special equipment and skills. On that matter, the authors have proposed a new practical technique to identify the absorption coefficient and dimensions of rooms. The technique comprises Subsystem_1 and Subsystem_2, each of which uses photographic images. Subsystem_1 uses a Gray Level Co-occurrence Matrix (GLCM) and is integrated with a Neural Network (NN) to identify the absorption coefficient of the material. Subsystem_2 uses a Dimension Vision Predictor (DVP) with the author's "ruler method" to identify the dimensions. Examinations conducted in practical rooms revealed a good correlation coefficient of $r \geq 0.90$ for Subsystem_1 and $r \geq 0.99$ for Subsystem_2. Finally, the System using NN gave inconsistent results, while FEA revealed consistent results with $r \geq 0.8$

Keywords: Neural network; gray level co-occurrence matrix (GLCM); photographic image; absorption coefficient; dimension; dimension vision predictor (DVP).

INTRODUCTION

Nowadays, various systems have been proposed to predict a room's acoustic properties, e.g. reverberation time (RT), and computational systems are quite popular in this field (e.g. ray tracing and numerical method) (Hodgson, 2009; Okuzono, 2010; Okuzono et al., 2012). There are two essential factors that make the RT values within appropriate ranges: i. the dimensions of the classroom, and ii. the absorption coefficients of materials. These kinds of factor are usually obtained from measurements, literature, references and so on, which requires extra time and expertise. An innovative system has been proposed by Hodgson (2005) to identify the absorption coefficients of materials in the classrooms using a statistical method (multivariable linear-regression techniques). Although it is a simple and fast system, it is still inadequate because it can only be used to determine the materials' absorption coefficients in classrooms at the University of British Columbia. Aiming for practical application, this study is to propose efficient systems using a photo imaging technique to predict the factors (as mentioned above) and to show their reliability when they are applied. In the subsystems, three techniques

are used: i. image processing (gray level co-occurrence matrix (GLCM)), ii. the Dimension Vision Predictor (DVP), and iii. a Neural Network (NN). Using combinations of the techniques, two systems are built: Subsystem_1 using GLCM and NN to determine the absorption coefficients of material surfaces in classrooms; and Subsystem_2 using DVP to determine the particular dimensions of rooms. The accuracy of the two systems is examined using actual rooms to investigate the identification capability of the two systems. The absorption coefficients and dimension identification from the two subsystems are used to compute RTs of actual rooms in System using a NN. The computed RTs are then compared with RTs by FEA using actual absorption coefficients and dimensions. Using these subsystems, we can ascertain room parameters easily, rapidly, and at a low cost compared to using physical measurement. The systems are useful for researchers, practical engineers, and designers to estimate the sound fields of existing rooms.

THEORETICAL DESCRIPTION

Gray Level Co-occurrence Matrix (GLCM)

The GLCM technique has been implemented successfully in texture feature analysis to analyze texture features of an image (Honeycutt & Plotnick, 2008). To date, no reported study has applied GLCM to acoustic fields. A GLCM is generated from a square matrix (N_g) with size determined according to the gray levels of pixels of an image that can be captured using a digital camera. An image includes numerous pixels, each of which presents a level of gray. A square matrix N_g is formed at these pixels. A GLCM comprises numerous elements, each designated as probability $P_{d,\theta}(i,j)$. The $P_{d,\theta}(i,j)$ represents pixels with gray levels i and j , which are counted at a certain distance d (e.g. $d = 1$ or 2) and direction angle θ ($\theta = 0^\circ, 45^\circ, 90^\circ$ and 135°) between the two image pixels. Haralick (Haralick et al., 1973) provides an additional explanation of GLCM. Figure 1 presents an example of computation of the GLCM with size $i = 3$ and $j = 3$. Here, i and j are taken from the gray level of an image. To count probability $P_{1,0^\circ}(3,0)$, by reference to Figure 1(a), it is three intensities of pair pixels ($i = 3, j = 0$) at a distance of two pixels $d = 1$; the direction angle θ is counted as 0° . A similar process can be conducted at $P_{1,90^\circ}(2,2)$. The intensity at that probability is 2. Generally, it is difficult to implement GLCM directly. Therefore, Haralick proposed 14 coefficients of texture features. The four commonly used Haralick coefficients are listed below.

$$cont = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P_{d,q}(i,j) \cdot (i-j)^2 \quad (1)$$

$$corr = \frac{\left\{ \hat{a}_{i=0}^{N_g-1} \hat{a}_{j=0}^{N_g-1} P_{d,q}(i,j) \times (i,j) \right\} - m_x m_y}{S_x S_y} \quad (2)$$

$$ASM = \hat{a}_{i=0}^{N_g-1} \hat{a}_{j=0}^{N_g-1} \{P_{d,q}(i,j)\}^2 \quad (3)$$

$$hom = \hat{a}_{i=0}^{N_g-1} \hat{a}_{j=0}^{N_g-1} \{P_{d,q}(i,j)\}^2 \quad (4)$$

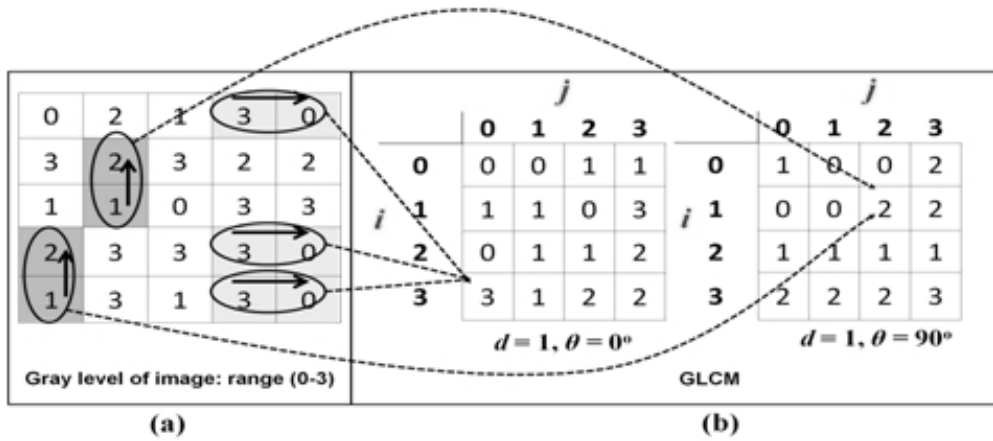


Figure 1. Computed GLCM: (a) gray level of an image, (b) GLCM for element $P_{1,0^\circ}$ (3,0) and at $P_{1,90^\circ}$ (2,2).

In those equations, *cont* is the contrast used to measure the image contrast, *corr* is the correlation used to measure image linearity, and *ASM* is the angular second moment used to measure image smoothness. Also, *hom* represents the homogeneity used to indicate homogeneity in uniform images. In addition, $\mu_x, \mu_y, \sigma_x,$ and σ_y are the respective means and standard deviations of the probability matrix of GLCM obtained by summing the row.

Dimension Vision Predictor (DVP)

Several techniques are used to measure dimensions using a camera. Some techniques demand special equipment and camera lens calibration. Therefore, aiming at practical use, this study chooses survey-from-photo because it can be implemented directly from any ordinary camera without calibrating the camera lens. Generally, survey-from-photo identifies the dimension based on two images. The images are marked with two corresponding points. Then both are connected to make a line at an object to measure. A reference dimension is necessary to achieve an accurate measurement. The reference dimension is a dimension obtained from an object that is known exactly. Here, survey-from-photo uses that object dimension as a reference to standardize the scale range to the images. The basic concept of survey-from-photo is that of the "stereo vision" principle, which uses two cameras to measure the dimensions of an object, as presented in Figure 2. One camera is located at Cr and another at Cl with intervening distance (d). The cameras are focused at points $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ with a certain focus length (f), which are all obtainable at the camera lens. At f , two image points are apparent at the image $P_{1r}, P_{1l}, P_{2r},$ and P_{2l} with respective coordinates $(x_{1r}, y_{1r}), (x_{1l}, y_{1l}), (x_{2r}, y_{2r}),$ and (x_{2l}, y_{2l}) . The coordinates $(x_{1r}, y_{1r}), (x_{1l}, y_{1l}), (x_{2r}, y_{2r}),$ and (x_{2l}, y_{2l}) are calculable by considering the center of the image as the origin. To obtain the coordinates of $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$, the equation is definable simply as shown:

$$P_1(x_1, y_1, z_1); \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \frac{d}{x_{1l} - x_{1r}} \begin{pmatrix} x_{1l} \\ x_{1r} \\ f \end{pmatrix} \tag{1}$$

$$P_2(x_2, y_2, z_2); \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \frac{d}{x_{2l} - x_{2r}} \begin{pmatrix} x_{2l} \\ x_{1l} \\ f \end{pmatrix} \quad (2)$$

$$L_x = d \frac{\frac{\partial}{\partial x_{1l}} x_{1l} - \frac{\partial}{\partial x_{2l}} x_{2l}}{\frac{\partial}{\partial x_{1l}} x_{1l} - \frac{\partial}{\partial x_{2l}} x_{2l}} \quad (3)$$

$$L_y = d \frac{\frac{\partial}{\partial x_{1l}} y_{1l} - \frac{\partial}{\partial x_{2l}} y_{2l}}{\frac{\partial}{\partial x_{1l}} x_{1l} - \frac{\partial}{\partial x_{2l}} x_{2l}} \quad (4)$$

$$L_z = d \frac{\frac{\partial}{\partial x_{2l}} 1 - \frac{\partial}{\partial x_{2r}} 1}{\frac{\partial}{\partial x_{2l}} x_{2l} - \frac{\partial}{\partial x_{2r}} x_{2r}} \quad (5)$$

The distance (L) between P_1 and P_2 can be simplified as:

$$L = \sqrt{L_x^2 + L_y^2 + L_z^2} \quad (6)$$

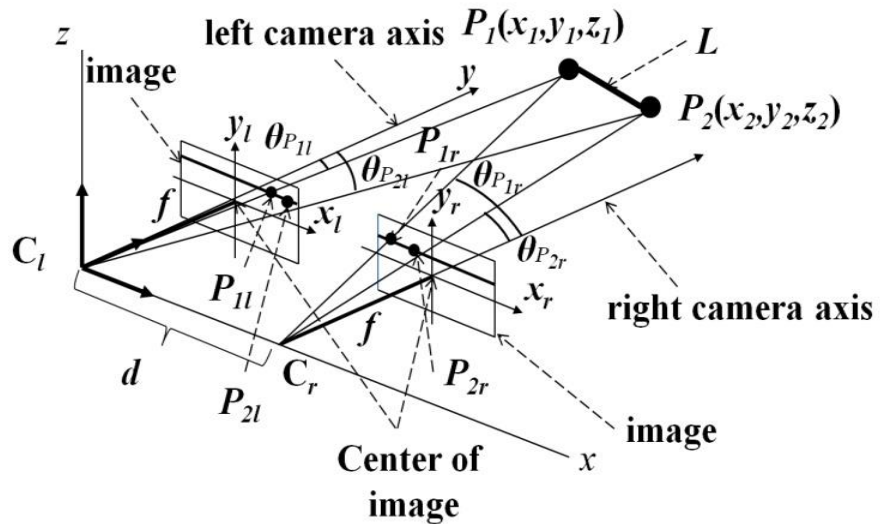


Figure 2. Principle of stereo vision.

Neural Network (NN)

Basically, NN architecture involves three layers: input layer, hidden layer and output layer. Each layer consists of a number of nodes to construct a network connection, as shown in Figure 3. Detailed explanation of NN can be found in a previous paper (Yahya et al., 2010). Before implementing the NN, a database to be analyzed is transformed

(0.1–0.9) to standardize the range. Overlearning occurs during the NN learning process. To surmount the problem, a database is divided into three subsets: a training subset, a validation subset, and a test subset. The training subset is used to train the NN. The validation subset is used to validate the learning process, and the test subset is used to investigate the prediction performance. The proportion of each subset is chosen randomly. To obtain the optimum network, 2–15 hidden nodes are used. The mean square error (MSE) and correlation coefficients (r) are used for assessment.

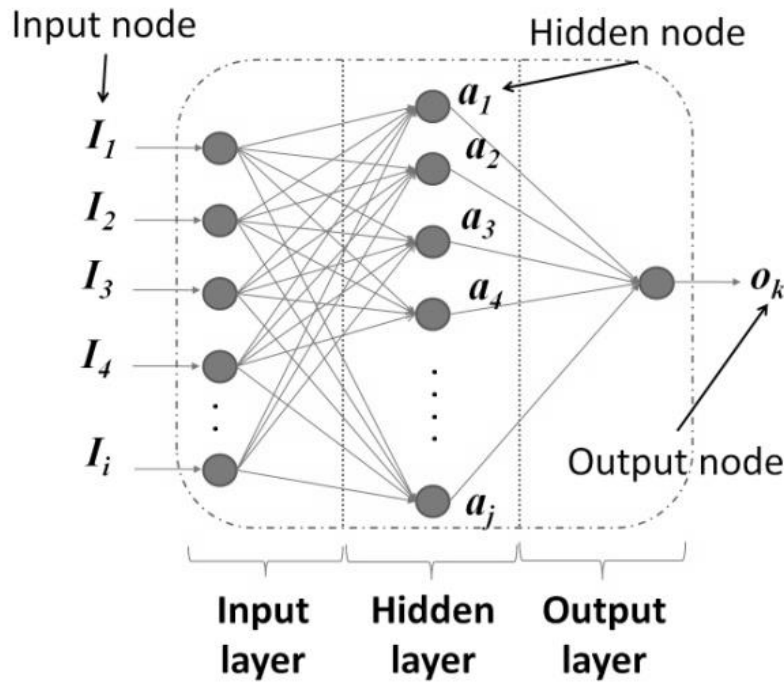


Figure 3. Architecture of NN

METHODOLOGY OF SUBSYSTEM_1

Material Surface Capturing

For this study, six material surfaces were taken of Oita University rooms, as portrayed in Figure 4. Surfaces (a), (b), (c), (d), (e), and (f) are, respectively, surfaces of walls, doors, floors, windows, ceilings, and carpets. To perform material surfaces capture, an ordinary camera is used. Regarding the standardization of images, a digital single-lens reflex (DSLR) camera with Sigma 50 mm f2.8 lens was used. The distance from the camera to the surface material was set to 40 mm with autofocus mode, while the respective lens settings for aperture, shutter speed and ISO speed were f2.8, 1/80, and 100. To analyze the accuracy of Subsystem_1, 368 images of surfaces were captured at different locations in three rooms. The proportions of images of material surfaces are: surface (a) = 69 images, surface (b) = 71 images, surface (c) = 66 images, surface (d) = 56 images, surface (e) = 67 images, and surface (f) = 40 images. All images were analyzed using GLCM.

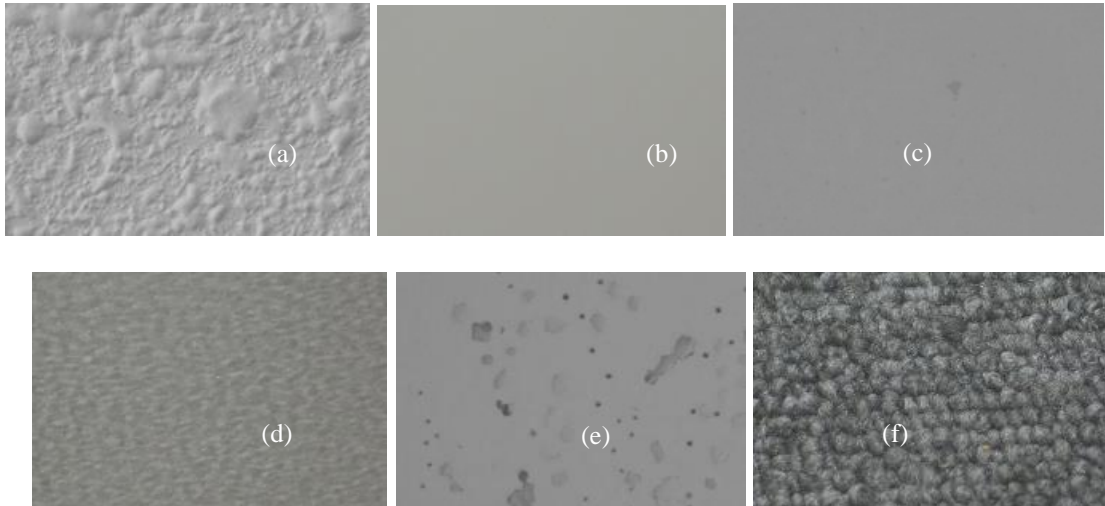


Figure 4. Sample images of material surfaces for (a) wall, (b) door, (c) floor, (d) window, (e) ceiling, and (f) carpet

GLCM Implementation

The GLCM was computed for the 368 images of surface materials using the following settings: i. $d=1, \theta=0^\circ$; ii. $d=1, \theta=45^\circ$; iii. $d=1, \theta=90^\circ$; and iv. $d=1, \theta=135^\circ$. Each Haralick's coefficient provides four values based on settings, but only an average value of the four values is considered hereinafter. The average value is designated as the coefficient value for this study. Because of variations of brightness and texture features in our experiment, the ranges of the coefficient values become too wide to be processed. To overcome this problem, a limitation for each coefficient value was made using the means (\bar{x}) and standard deviation (σ). The limitations are $(\bar{x}-\sigma)$ and $(\bar{x}+\sigma)$, respectively, for low limitation and high limitation. The coefficient values beyond the limitations were removed from further investigation.

FFNN Implementation

Coefficient values in the limitation were fed into the NN. Four coefficients (*cont*, *corr*, *ASM*, and *hom*) and the material surface were used respectively as input nodes and output nodes. Then the numbers of hidden nodes were set up as described previously. The learning algorithm chosen was Levenberg–Marquardt (*trainlm*) because it is faster and more efficient. To obtain the optimum network, a trial and error scheme was conducted by combining all those nodes (e.g. [i; h; o] for [input node; hidden node; output node]; example combination [4, 6, 1], [4,10,1], ... or [4, 9, 1]) but only one combination that provided good performance was selected.

METHODOLOGY OF SUBSYSTEM_2

DVP Implementation

The same camera as for Subsystem_1 was used with focus lenses of 18–70 mm to capture two images at one view. The camera was set in autofocus mode. Figure 5

presents an example of predicting dimensions of objects in one image at one view. Lines connect corresponding points at objects. For example, to measure the dimensions of a blackboard, four corresponding points of A, B, C, and D must be obtained. Each corresponding point is connected to form lines: line 9, 10, 11, and 12. As described above, the survey-from-photo requires a standard scale. Therefore, the authors propose to use a ruler that is attached at an appropriate view as a reference dimension in this "ruler method". A ruler is preferred because it is practical and simple to attach to the view region to be measured. To investigate the repeatability of dimension prediction, 100 dimensions of several objects were examined. The predicted dimensions using DVP as Subsystem_2 were compared with measured dimensions obtained from laser measurements using a laser indicator (LS-501A; MAX Co., Ltd.). The MSE and correlation coefficient (r) are applied for assessments.

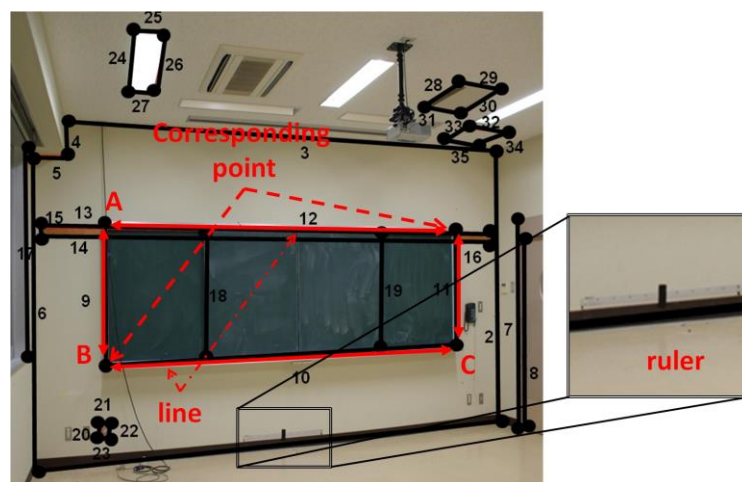


Figure 5. Dimension identification.

METHODOLOGY OF SYSTEM

FEA Implementation

To obtain the RTs database of rooms which are used for construction of the NN, 20 rooms with different volumes were simulated using Finite Element Analysis (FEA). In the simulation, six absorption coefficients of surface materials of the wall, door, window, floor, ceiling and furniture were considered in this study. Basically, the absorption coefficient values (ranging from 0 to 1) depend on the type of material, either reflective or absorptive. Considering all of the absorption coefficient values will increase the computing time and cost of FEA. To overcome the problem, two kinds of condition were considered; i. dead: ($\alpha_w = 0.08$; $\alpha_{dr} = 0.1$; $\alpha_{wdw} = 0.4$; $\alpha_{flr} = 0.06$; $\alpha_{clg} = 0.4$; $\alpha_f = 0.4$); ii. live: ($\alpha_w = 0.02$; $\alpha_{dr} = 0.02$; $\alpha_{wdw} = 0.04$; $\alpha_{flr} = 0.02$; $\alpha_{clg} = 0.2$; $\alpha_f = 0.4$), where α_w , α_{dr} , α_{wdw} , α_{flr} , α_{clg} , and α_f represent the absorption coefficients for the wall, door, window, floor, ceiling and furniture, respectively. Dead is the maximum value of the absorption coefficient, while live is the minimum value of the absorption coefficient. These conditions were obtained from several surface materials in the Oita University room. Furthermore, another 6 rooms were simulated using FEA. These rooms were used to test the performance of the NN.

NN Implementation

The database of 1220 RTs obtained from 20 simulated rooms by FEA were fed into the NN. The database was divided into two subsets, the training subset (70% of database) and validation subset (30% of database). To confirm the reliability of the prediction, the testing database of 360 RTs obtained from the six simulated rooms was involved.

RESULTS AND DISCUSSION

Subsystem_1

Figure 6 shows the range of four coefficient values for six materials. Each coefficient value portrays a limitation. The limitations indicated that only 53.8% of 360 images of surfaces (surface (a) = 39 images, surface (b) = 50 images, surface (c) = 27 images, surface (d) = 26 images, surface (e) = 30 images, and surface (f) = 20 images) were used for the NN as input nodes because of limitations. Before feeding into the NN, a database of images of surface materials was divided into three subsets: 60% of the database for training; 20% of the database for validation, and 20% of the database for testing. No specific proportions for the NN subsets were set. At this point, the proportions of subsets are chosen arbitrarily. Generally, the training subset should be larger than the validation subset and the testing subset.

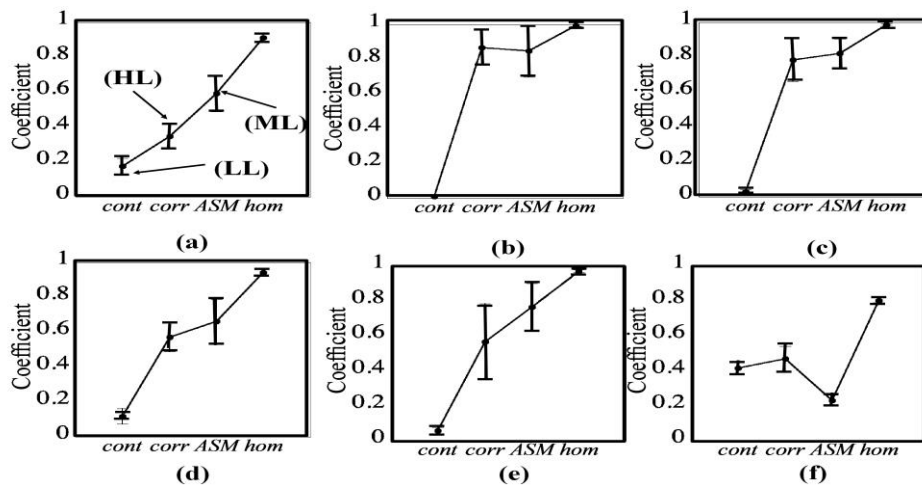


Figure 6. Limitation of coefficient values for six materials.

In this study, the absorption coefficients (α) of six material surfaces are derived from reports in the relevant literature (Maekawa & Lord, 1993). By identifying the material surfaces, we are able to ascertain the absorption coefficients of surfaces simultaneously. To identify the material surfaces, we used a classification number (1–5) to represent the output parameter: 1. Surface (a) ($\alpha = 0.07$), 2. Surface (b) and (c) ($\alpha = 0.02$), 3. Surface (d) ($\alpha = 0.04$), 4. Surface (e) ($\alpha = 0.4$) and 5. Surface (f) ($\alpha = 0.06$). Results of analyses show that the optimum network [4, 6, 1] with $MSE \leq 0.0018$ and $r \geq 0.9$ was obtained for both the training and validation subsets. To confirm their performance, the testing subset (39 surface images) showed $MSE \leq 0.07$ with $r \geq 0.9$. Subsystem_1 performance is inferred to be good at this stage. The restrictions of Subsystem_1 are the following: 1. It can only identify the material surfaces depending

on the database of material surfaces used. If more databases of material surfaces are used, then more material surfaces can be identified. 2. Generally, the real absorption coefficients of material surfaces in rooms depend on the material thickness, the presence or absence of an air layer and absorptive layer, and so on. Then, it is difficult to obtain a real absorption coefficient from only a surface form image. For practical usage, the author referred to related reports of the absorption coefficient.

Subsystem_2

The results of analyses of Subsystem_2 are given in Figure 7, which reveals a high correlation coefficient ($r \geq 0.99$) between the predicted values using Subsystem_2 and the measured values, with $MSE \leq 0.009$. The results show that Subsystem_2 provided high reliability using no physical measurements.

System

From the analysis, the optimum network is [4,11,1], for which the training and validation database indicate $MSE \leq 0.0012$ with $r \geq 0.9$ for both of them. Furthermore, for confirmation the testing database indicated $MSE \leq 0.007$ and $r \geq 0.80$. At this point, the NN used for the System is developed. It gives good reliability of prediction RTs on the six simulated rooms.

Implementation in Actual Room

Four types of actual rooms were utilized to investigate the prediction reliability of the subsystems and System. At Subsystem_1, the 294 surface images (surface (a) = 60 surface images, surface (b) = 53 surface images, surface (c) = 48 surface images, surface (d) = 63 surface images, surface (e) = 41 surface images and surface (f) = 25 surface images) were captured. However, only 180 surface images were selected after normalization (\bar{x}). These surface images were fed into the NN. At Subsystem_2, the target objects were the room, door, window and furniture (desk and chair). These objects were captured and fed into the DVP to predict the dimensions. Later on, predictions from both Subsystems were moved to the System using the NN. Besides that, the predictions from Subsystem_1 and Subsystem_2 were also moved to FEA. The prediction reliability of RT using System and FEA were examined. Figure 8 reveals the prediction by Subsystem_1, which gives high $r \geq 0.9$. Unfortunately, predictions on three (Surface (d): window ($\alpha = 0.04$)) showed inconsistent results. This is because 29% of the 31 surface images were indicated as below the limit at ASM. The window is a transparent and light-reflecting material. In the case of this transparent window, it was difficult to capture consistent surface images due to the nature of the material surface.

Figure 9 shows the prediction results of the dimensions of the room. Following the capturing procedure, the DVP produces high $r \geq 0.9$ in predicting the dimensions. To predict RTs, the System (using NN) and FEA used the same input parameters obtained from Subsystem_1 and Subsystem_2. The System provided predictions that were inconsistent with the FEA. On the other hand, by using the actual input parameter obtained from the four rooms, the FEA gave a consistent prediction. Figure 10 shows that the correlation between FEA actual and FEA predicted is more than 0.80. From the observation, at the moment, the techniques in Subsystem_1 and Subsystem_2 provide good prediction reliability when they are utilized with FEA.

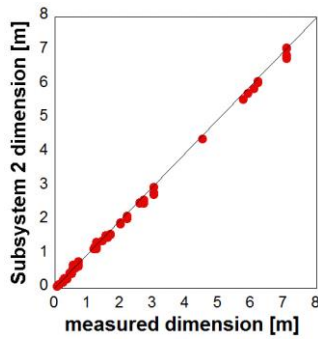


Figure 7. Correlation for Subsystem_2

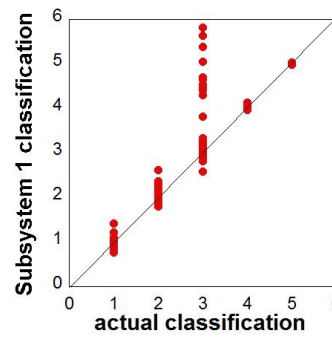


Figure 8. Correlation for Subsystem_1 (actual room)

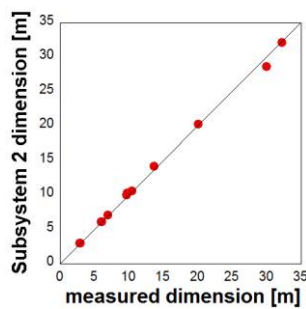


Figure 9. Correlation for Subsystem_2 (actual room)

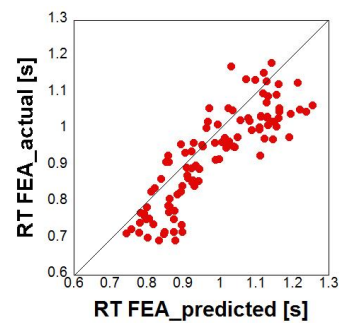


Figure 10. Correlation between FEA actual and FEA predicted

CONCLUSION

Aiming for practical application, a technique for predicting the absorption coefficient and dimensions using photographic images and an NN was developed. In Subsystem_1, a GLCM and NN are used to predict the absorption coefficient, while in Subsystem_2 DVP is used to predict the dimensions. The results from Subsystem_1 and Subsystem_2 show good reliability, with $r > 0.9$. The System and FEA are used to predict RTs. Comparison between them shows that FEA offers more consistent results, with $r \geq 0.8$. From the overall results, we can conclude that by applying Subsystem_1 and Subsystem_2, the practical value of the absorption coefficient and dimensions can be predicted. At this stage, the predicted values from Subsystem_1 and Subsystem_2 are useful to predict the RTs of room by using consistent methods such as FEA, BEM, and empirical methods.

ACKNOWLEDGMENT

This study is supported by RAGS (KPT) grant – R014, managed by the Research, Innovation, Commercialization and Consultancy Office (ORICC), UTHM.

REFERENCES

- Haralick, R. M, Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 6, 610-621.
- Hodgson, M. (2009). Ray-tracing prediction of optimal conditions for speech in realistic classrooms. *Applied Acoustics*, 70, 915-920.
- Hodgson, M. (2005). Estimate of the absorption coefficients of the surfaces of classrooms. *Applied Acoustics*, 67, 936-944.
- Honeycutt, C. E. & Plotnick, R. (2008). Image analysis techniques and gray-level co-occurrence matrices (GLCM) for calculating bioturbation indices and characteristic biogenic sedimentary structure. *Computer & Geoscience*, 34, 1461-1472.
- Maekawa, Z. & Lord, P. (1993). *Environmental and architectural acoustics*. London: Spon Press.
- Okuzono, T. (2010). Fundamental accuracy of time domain finite element method for sound-field analysis of rooms. *Applied Acoustics*, 71(10), 1027-1067.
- Okuzono, T., Otsuru, T., Tomiku, R., & Okamoto, N. (2012). Application of modified integration rule to time-domain finite element acoustics simulation of rooms. *Journal of Acoustical Society of America*, 132(2), 804-813.
- Yahya, M. N., Otsuru, T., Tomiku, R., & Okuzono, T. (2010). Investigation the capability of neural network in predicting reverberation time on classroom. *International Journal of Sustainable Construction Engineering & Technology*, 19-31.