Type/position classification of inter-floor noise in residential buildings with a single microphone via supervised learning

Hwiyong Choi, Haesang Yang, Seungjun Lee, and Woojae Seong*
Seoul National University, Seoul, Korea

***Abstract*—Inter-floor noise propagates through a building structure from a noise source to neighbors on other floors. Identification of type/position of inter-floor noise in a building is difficult for human hearing. A convolutional neural network-based inter-floor noise type/position classification method was proposed in [*Appl. Sci.* 9, 3735 (2019)] to identify inter-floor noise. The method was evaluated against inter-floor noise collected in a single campus building as a feasibility test. In this work, the generalizability of the method was addressed through numerous tasks using new datasets collected in two real apartment buildings. These datasets contain inter-floor noise generated in rooms and at positions with three-dimensional spatial diversity, which was not studied in the previous work. Furthermore, type classification knowledge transfer between two individual apartment building domains was studied.**

***Index Terms*—Inter-floor noise, single sensor acoustics, convolutional neural network, knowledge transfer**

I. Introduction

In a multi-dwelling unit, inter-floor noise from a noise source propagates through slabs, columns, walls, air, and arrives at the human ears. The arrival of this noise on other floors can annoy residents [1]. This inter-floor noise is a serious problem in a city where most of the residential buildings are multi-dwelling units [2]. For example, the Floor Noise Neighborhood Centre of the Korea Environment Corporation received 137,813 complaints of inter-floor noise during the year 2012–2018 [3]. Identification of type/position of an inter-floor noise is difficult for human hearing. Sometimes, misjudgement of noise type/position causes a conflict between neighbors. From this point of view, an inter-floor noise type/position classifier can reduce such misjudgement and prevent the conflict among neighbors [4].

Indoor occupant localization is a related area of research on inter-floor noise classification. Localization of footsteps induced by a person in a building was studied to understand human behavior and for building energy management [5]–[7]. These techniques localize a footstep on a slab via estimation of TDOA (time difference of arrivals) using signals measured with seismic sensors or accelerometers.

In the authors' previous works [4], [9], a supervised learning-based inter-floor noise classifier was proposed and the feasibility of the method was shown against a real-world dataset. This approach assumed a measured inter-floor noise signal as a response to a given impact on a floor. Inter-floor noise signals were sampled with a single microphone and converted to log-scaled Mel-spectrograms. And the log-scaled Mel-spectrograms were classified into inter-floor noise type/position categories using a CNN (convolutional neural network)-based model. The contributions of the previous work are the classification of inter-floor noise on the floors above/below utilizing the generalized knowledge via transfer learning. This method has an advantage in mobile application development, because it uses a single sensor. Whereas the researches on indoor occupant localization focused on localization of footstep noises on the same floor using signals over multiple channels of sensors [5]–[7].

Although the previous work demonstrated the feasibility of the method, it was shown with a dataset collected on a corridor in a single campus building. Also, inter-floor noise positions were nearly in two-dimension. Therefore, the generalizability needs to be demonstrated for real-world application in residential buildings with more complex spatial diversity. Furthermore, the viability of classification knowledge transfer between similar individual building domains can reduce the effort for site surveying and dataset gathering.

The contributions of this paper can be summarized as follows. (1) Inter-floor noise events were collected in two different apartment buildings. These new datasets contain inter-floor noise generated in rooms and at positions with three-dimensional spatial diversity, which potentially make this work closer to real-world application scenarios. (2) Numerous inter-floor noise classification tasks considering real-world applications were prepared and tested. Also, the limitations of this approach were discussed. (3) A type classification model trained on inter-floor noise in one building was tested against a dataset in the other building to demonstrate the viability of type classification knowledge transfer.

II. Inter-floor Noise Dataset

Inter-floor noise signals were newly collected in two apartment buildings [8] to evaluate the performance of the inter-floor noise classifier against inter-floor noises in real residential buildings.

Noise types and source positions were selected based on a report [3], which provides the identified inter-floor noise

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2019R1F1A1058794).
*Electronic mail: wseong@snu.ac.kr

types and positions. The report shows that the main identified inter-floor noise types are footsteps, hammering, furniture, home appliances, and doors. And 95.7% of the identified noise sources are on the floors above/below.

The selected noise types are: a medicine ball falling to the floor from a height of 1.2 m (MB), a hammer dropped from 1.2 m above the floor (HD), hammering (HH), dragging a chair (CD), and running a vacuum cleaner (VC). They were generated on the floors above/below based on the position of receiver in the apartment buildings and sampled over a single microphone in a smartphone (Samsung, Galaxy S6) with sampling frequency f_s of 44,100 Hz. These new datasets are extensions to the authors' previous work [4], which contains the same noise types.

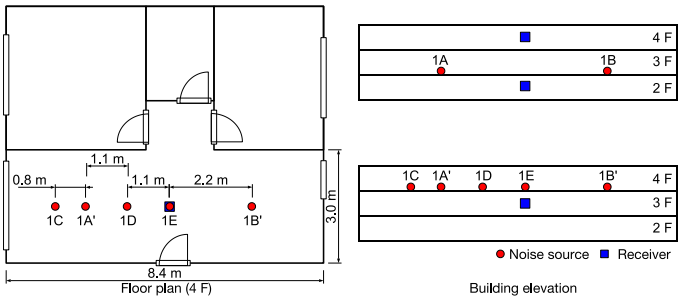
Fig. 1 shows floor plans and building elevations of the two apartment buildings. The apartment buildings are reinforced concrete frame structure and partitioned with concrete walls. The floors are covered with vinyl floorings. The circles and the squares in the figure are noise sources and receiver positions, respectively.

The inter-floor noise signals were generated and collected as follows: (1) In apartment building 1 (APT 1), MB, HD, HH, CD, and VC were generated at 1-A and 1-B on 3 F and sampled using the single receiver on the floors above/below as a training/validation dataset; (2) The same noise types were generated at 1-A′ and 1-B′ on the floor above (4 F) and sampled with the single receiver on the floor below as a test dataset. MB and HH were generated at 1C, 1D, and 1E to test the robustness of a classifier to a positional variation on the same floor; (3) In apartment building 2 (APT 2), MB, HD, HH, and CD were generated at 2A, 2B, 2C, and 2D on 4 F as a training/validation dataset. They were sampled with the single receiver on the floors above/below; and (4) The same noise types were generated at 2A′, 2B′, 2C′, and 2D′ as a test dataset. VC was not generated because of the electricity cut off in APT 2 during dataset gathering.

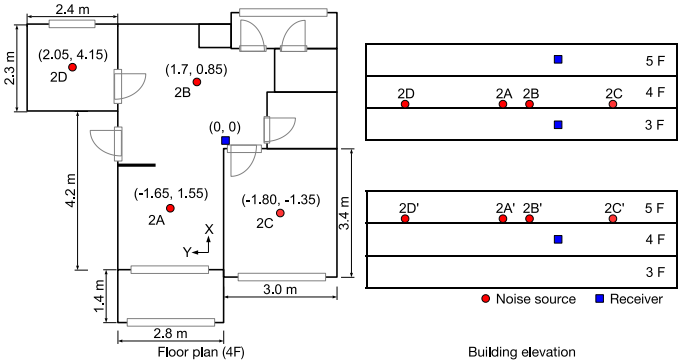
During the data collection in APT 1, the only allowed experimental site was the living rooms. For this reason, the dataset contains inter-floor noises generated only in the living rooms. Whereas APT 2 dataset contains inter-floor noises generated in the bed rooms as well as in the living rooms. The number of data collected in APT 1 and APT 2 are 1,785 and 2,880, respectively. An event start of each inter-floor noise sample was visually annotated.

III. INTER-FLOOR NOISE CLASSIFICATION

In this section, an inter-floor noise representation $\mathbf{I}$ is classified into an inter-floor noise type category y_{t} or a position category y_{p} by classifiers $f_{\mathbf{W}_{\mathrm{t}}}(\cdot)$ or $f_{\mathbf{W}_{\mathrm{p}}}(\cdot)$, where $\mathbf{W}_{\mathrm{t}}$ and $\mathbf{W}_{\mathrm{p}}$ represent weights of the type and position classifier, respectively. And diversified tasks are designed and tested against to the datasets. All the designed tasks were implemented and validated [8] with TensorFlow [10].



(a) Apartment building 1 (APT 1)



(b) Apartment building 2 (APT 2)

Fig. 1. Floor plans with XY positions of the noise sources and receivers and building elevations.

A. Input representation

Log-scaled Mel-spectrogram $\mathbf{P} \in \mathbb{R}^{H\times W}$ was used as $\mathbf{I}$, where H and W are defined by the size of the CNN input layer. $\mathbf{P}$ was obtained as follows.

For a given inter-floor noise signal $\mathbf{s}(q) \in \mathbb{R}^{l_s}$ with amplitudes between $[-1,\ 1]$ and l_s =132,300 (3 in seconds), a short time samples with size of l_w =1,024 to have spectral line resolution of 43 Hz is windowed (Hanning window) and zero padded by $\mathrm{ZP}_N(\cdot)$:

$$\begin{aligned}\mathbf{x}_w = \mathrm{ZP}_N\left(\mathbf{s}\left(q\right)\right) \in \mathbb{R}^N, \quad hw \le q \le hw + l_w - 1,\\ w = 0, 1, \cdots, W-1, \quad N = 8{,}192.\end{aligned} \tag{1}$$

l_w allows $\mathbf{s}(q)$ to contain low frequency band. The next short time samples are selected using hop size of $h = 591$ which yields $H = W = 224$.

$\mathbf{x} = [\mathbf{x}_0^{\mathsf{T}} \mid \cdots \mid \mathbf{x}_{W-1}^{\mathsf{T}}] \in \mathbb{R}^{N\times W}$ is converted to a power spectrogram

$$\begin{aligned}\mathbf{X}_w(k) = |\mathrm{DFT}[\mathbf{x}_w(n)]|^2 = \left|\sum_{n=0}^{N-1} \mathbf{x}_w(n) e^{-j2\pi nk/N}\right|^2,\\ w = 0, 1, \cdots, W-1, \qquad k = 0, 1, \cdots, N-1.\end{aligned} \tag{2}$$

A Mel-spectrogram $\mathbf{M}$ is obtained as $\mathbf{M} = \mathbf{F}\mathbf{X}$, where the Mel-filterbank $\mathbf{F}$ changes the frequency scale of $\mathbf{X}$ to the Mel-scale. Since almost of the signals in the datasets have power

below 2 kHz, the maximum frequency of $\mathbf{F}$ is set as 2 kHz. And the scale of each entry of $\mathbf{M}$ is changed

$$\mathbf{P} = 10\ \log(\mathbf{M}/\max\ \mathbf{M}_{i,j}). \tag{3}$$

B. Inter-floor noise classifier

In this work, VGG16 [11] with an adaptation layer was employed as $f_{\mathbf{W}}(\cdot)$. $\mathbf{W}$ was initialized with weights from VGG16 trained for recognition of images in ImageNet [12]. This approach, called the transfer learning [13], [14], demonstrated the effectiveness in inter-floor noise classification with data sparsity [9]. Other CNN architectures designed for image recognition can be used for inter-floor noise type/position classification [4]. Among the CNNs employed in the previous study, VGG16 was selected and used in this work, which showed the best performance.

$f_{\mathbf{W}}(\cdot)$ was optimized via minimizing a loss

$$\begin{gathered}\mathcal{L}\left(\mathbf{W}, \lambda\right) = -\sum_{i=1}^{C} \mathbf{y}_i \log \hat{\mathbf{y}}_i + \lambda \left\|\mathbf{w}\right\|_2^2, \\ \text{where } \hat{\mathbf{y}}_i = \operatorname{softmax}\left(f_{\mathbf{W}}(\mathbf{P})\right)\end{gathered} \tag{4}$$

and $\mathbf{y}_i$ represents a one-hot-encoded category C of a given $\mathbf{P}$. The loss is a sum of a cross entropy loss and a regularization of weights of the adaptation layer. $\mathcal{L}$ was minimized through mini-batch gradient descent with batch size of 64 for 50 epochs. The learning rate η and the regularization strength λ were found using the random search [15].

IV. Tasks

Tasks were prepared as shown in Fig. 2 to evaluate $f_{\mathbf{W}}(\cdot)$ against the new inter-floor noise datasets collected in APT 1 and APT 2: (1) Type classification of inter-floor noises at learned and unlearned positions; (2) Position classification of inter-floor noises at learned and unlearned positions; and (3) Inter-floor noise type classification knowledge transfer between APT 1 and APT 2 domains. Mean cross-validation accuracy or mean test accuracy (%) were measured for evaluation of each task.

A. Inter-floor noise type classification tasks

The following five-fold cross-validation tasks were prepared to address type classification of inter-floor noise generated on the floors above/below in the two individual apartment buildings, where the name of each task represents [apartment building]-[type classification]-[cross-validation].

APT1-T-CV [(a) in Fig. 2]. The inter-floor noise signals at 1A and 1B in APT 1 with noise type labels $y_{\mathrm{t}} \in \mathcal{Y}_{\mathrm{t,APT\ 1}}$ are converted to $\mathbf{P}$. A standard five-fold cross-validation is performed against $\mathbf{P}$ with $y_{\mathrm{t}} \in \mathcal{Y}_{\mathrm{t,APT\ 1}}$.

APT2-T-CV [(b) in Fig. 2]. The inter-floor noise signals at 2A, 2B, 2C, and 2D in APT 2 with noise type labels are converted to $\mathbf{P}$. A standard five-fold cross-validation is performed against $\mathbf{P}$ with $y_{\mathrm{t}} \in \mathcal{Y}_{\mathrm{t,APT\ 2}}$.

The following two tasks test the $f_{\mathbf{W}_{\mathrm{t}}^*}(\cdot)$ against signals from unlearned positions, where the name of each task represents [apartment building]-[type classification]-[test].

APT1-T-T [(c) in Fig. 2]. The type classifier with the optimal hyperparameters $f_{\mathbf{W}_{\mathrm{t},j}^*}(\cdot)$, $(j = 1, \cdots, 5)$ obtained in APT1-T-CV is tested against the inter-floor noise signals from 1A′, 1B′, 1C, 1D, and 1E in APT 1.

APT2-T-T [(d) in Fig. 2]. $f_{\mathbf{W}_{\mathrm{t},j}^*}(\cdot)$ obtained in APT2-T-CV is tested against the inter-floor noise signals from 2A′, 2B′, 2C′, and 2D′ in APT 2.

B. Inter-floor noise position classification tasks

The following five-fold cross-validation tasks address the noise source position classification within individual apartment buildings, where the name of each task represents [apartment building]-[position classification]-[cross-validation].

APT1-P-CV [(e) in Fig. 2]. The inter-floor noise signals in form of $\mathbf{P}$ from 1A and 1B in APT 1 are labeled as their positions y_{p} relative to the receiver positions

$$\{\text{1A-a, 1A-b, 1B-a, 1B-b}\} \in \mathcal{Y}_{\mathrm{p,APT\ 1}}, \tag{5}$$

where a and b represent the floors above and below, respectively. The standard five-fold cross-validation is performed against $\{\mathbf{P},\ y_{\mathrm{p}}\}$, $y_{\mathrm{p}} \in \mathcal{Y}_{\mathrm{p,APT\ 1}}$.

APT2-P-CV [(f) in Fig. 2]. This task performs five-fold cross-validation against inter-floor noise signals in form of $\mathbf{P}$ in APT 2 with position labels y_{p}

$$\begin{gathered}\{\text{2A-a, 2A-b, 2B-a, 2B-b, 2C-a, 2C-b,} \\ \text{2D-a, 2D-b}\} \in \mathcal{Y}_{\mathrm{p,APT\ 2}}.\end{gathered} \tag{6}$$

The following two tasks test the $f_{\mathbf{W}_{\mathrm{p}}^*}(\cdot)$ against the inter-floor noise signals in form of $\mathbf{P}$ from the unlearned positions within individual apartment buildings. The name of each task represents [apartment building]-[position classification tasks]-[test].

APT1-P-T [(g) in Fig. 2]. $f_{\mathbf{W}_{\mathbf{p,APT\ 1}}^*}(\cdot)$ obtained in APT1-P-CV is tested whether it maps $\mathbf{P}$ with 1A′-a to 1A-a and that with 1B′-a to 1B-a. The subscript $\mathrm{p,APT\ 1}$ denotes $\mathbf{W}^*$ with the optimal hyperparameters is trained/validated against $\{\mathbf{P},\ y_{\mathrm{p}}\}$, $y_{\mathrm{p}} \in \mathcal{Y}_{\mathrm{p,APT\ 1}}$. If the classifier works well on this task, data gathering for all floors is not required. Also, $f_{\mathbf{W}_{\mathbf{p,APT\ 1}}^*}(\cdot)$ is tested against $\mathbf{P}$ with 1C-a, 1D-a and 1E-a. These categories are totally new to the classifier. So this can address viability of the approach for finding a floor where an inter-floor is generated. It is considered to be the main concern in this study and close to real situation.

APT2-P-T [(h) in Fig. 2]. $f_{\mathbf{W}_{\mathbf{p,APT2}}^*}(\cdot)$ obtained in APT2-P-CV is tested against $\mathbf{P}$ with 2A′-a, 2B′-a, 2C′-a, and 2D′-a.

C. Inter-floor noise type classification knowledge transfer

Two tasks were prepared to demonstrate type classification knowledge transfer between two apartment buildings.

APT2|APT1 [(i) in Fig. 2]. $f_{\mathbf{W}_{\mathbf{t,APT1}}^*}(\cdot)$ is tested against all $\mathbf{P}$ in APT 2 and results were observed.

APT1|APT2 [(j) in Fig. 2]. $f_{\mathbf{W}_{\mathbf{t,APT2}}^*}(\cdot)$ is tested against all $\mathbf{P}$ in APT 1.

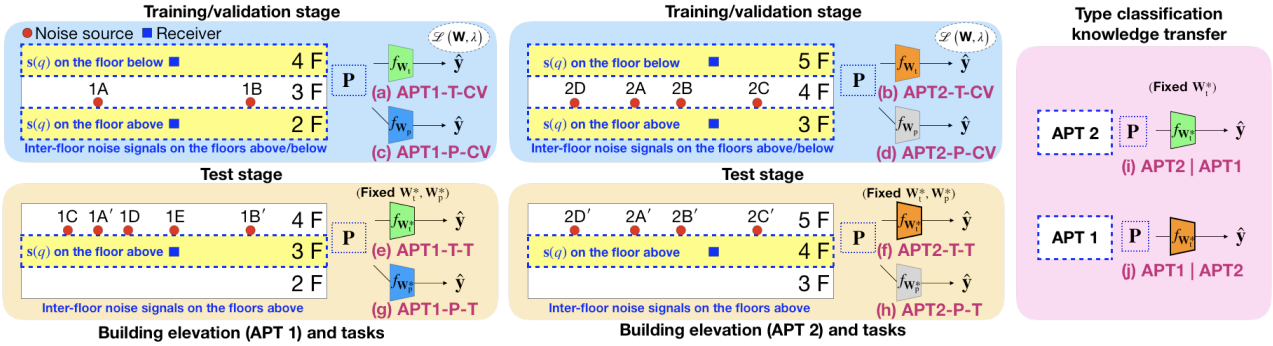


Fig. 2. Inter-floor noise classification tasks. (a)–(j) denote the tasks as explained in Sections IV-A–IV-C.

V. RESULTS AND DISCUSSIONS

A. Inter-floor noise type classification results

Table I shows results of the type classification tasks as described in Section IV-A. These results were arranged by noise types.

For the cross-validation tasks, the method showed average accuracies of 97.5% and 94.3%, respectively. Almost of the misclassification in APT2-T-CV were the confusions of HD with HH. These confusions could be explained, because the both noise types were generated by a hammer.

For the test tasks, the method showed average accuracies of 84.9% and 91.7%, respectively. The signals from the unlearned positions yielded the lower average accuracies than those of the cross-validation tasks. Nonetheless, these results showed the feasibility of the type classification approach against inter-floor noise signals transmitted through unlearned floor sections. Potentially, we can reduce the human effort for data gathering when this approach is employed.

B. Inter-floor noise position classification results

Table II and IV show confusion matrices drawn with results of APT1-P-CV and APT2-P-CV, respectively. Their average accuracies were 96.1% and 93.8%, respectively. The most confused categories in the both tasks were positions on the same floor *e.g.*, 1A-b and 1B-b.

Table III and V show confusion matrices drawn with results of APT1-P-T and APT2-P-T, respectively. As shown in the tables, the method showed its limitation of noise position classifications from unlearned positions. For example, 40.6% of 1A′-a was mapped to 1B-a as shown in Table III.

Nonetheless, the method showed a strong prediction performance of a signal from an unlearned position. Also, it can predict floor of a noise signal from an unlearned position around the known positions, *e.g.*, signals with 1C-a, 1D-a, and 1E-a were generated at unlearned positions but at around 1A′-a and 1B′-a whose two-dimensional positions are same as 1A-a and 1B-a. The measured floor classification accuracies of APT1-P-T and APT2-P-T were 98.0% and 96.4%, respectively. In a real-world situation, inter-floor noise does not always occur at a learned position. The result of APT1-P-T showed the proposed method can deal with such positional variability.

C. Inter-floor noise type classification knowledge transferring results

The results of APT2|APT1 and APT1|APT2 were drawn into confusion matrices as in Table VI and VII. Confusions between HD and HH categories were observed. And the results showed that $f_{\mathbf{W}_{t,\mathrm{APT\ 2}}}(\cdot)$ generalizes noise type knowledge better than $f_{\mathbf{W}_{t,\mathrm{APT\ 1}}}(\cdot)$. A reason for this could be found in the datasets. The number of data in APT 2 is 1.6 times larger than that of APT 1. Inter-floor noise signals in APT 1 were generated only in the living rooms, whereas APT 2 contains samples generated in the bed rooms as well as in the living room covering wider range. Potentially $f_{\mathbf{W}_{t,\mathrm{APT\ 2}}}(\cdot)$ could learn various cases with a larger number of data.

TABLE I
TYPE CLASSIFICATION ACCURACIES (%)

	MB	HD	HH	CD	VC	Average
APT1-T-CV	97.4	96.0	99.0	98.5	96.0	97.5
APT2-T-CV	99.0	87.1	91.3	100.0	-	94.3
APT1-T-T	94.0	65.7	99.2	97.6	78.2	84.9
APT2-T-T	98.3	96.7	73.2	98.8	-	91.7

TABLE II
CONFUSION MATRIX FOR APT1-P-CV RESULT (%)

		Predicted Label			
		1A-a	**1B-a**	**1A-b**	**1B-b**
True Label	**1A-a**	**98.4**	0.4	0.4	0.8
	1B-a	0.8	**98.0**	0.4	0.8
	1A-b	0.0	0.0	**95.5**	4.5
	1B-b	0.0	0.0	8.9	**91.1**

TABLE III
CONFUSION MATRIX FOR APT1-P-T RESULT (%)

		Predicted Label			
		1A-a	**1B-a**	**1A-b**	**1B-b**
True Label	**1A′-a**	**57.4**	40.6	0.7	1.3
	1B′-a	42.8	**54.4**	0.6	2.2
	1C-a	**57.7**	42.3	0.0	0.0
	1D-a	36.4	**62.4**	0.2	1.0
	1E-a	48.2	**48.8**	0.8	2.2

TABLE IV
CONFUSION MATRIX FOR APT2-P-CV RESULT (%)

		Predicted Label							
		2A-a	**2B-a**	**2C-a**	**2D-a**	**2A-b**	**2B-b**	**2C-b**	**2D-b**
True Label	**2A-a**	**96.7**	0.0	2.1	0.8	0.0	0.0	0.4	0.0
	2B-a	0.4	**92.1**	5.0	1.3	0.0	0.0	0.4	0.8
	2C-a	0.4	2.9	**90.4**	2.9	0.8	0.8	1.3	0.4
	2D-a	0.4	0.4	2.1	**94.6**	0.0	0.0	0.8	1.7
	2A-b	0.0	0.0	0.0	0.8	**97.1**	1.3	0.4	0.4
	2B-b	0.0	0.8	0.0	0.0	1.3	**91.7**	5.8	0.4
	2C-b	0.0	1.3	2.1	0.0	0.4	2.1	**92.9**	1.3
	2D-b	0.0	0.0	0.4	2.1	0.0	0.4	2.1	**95.0**

TABLE V
CONFUSION MATRIX FOR APT2-P-T RESULT (%)

		Predicted Label							
		2A-a	**2B-a**	**2C-a**	**2D-a**	**2A-b**	**2B-b**	**2C-b**	**2D-b**
True Label	**2A′-a**	27.8	18.8	6.9	**40.3**	0.0	0.0	0.1	6.1
	2B′-a	5.8	**55.1**	11.8	26.8	0.0	0.1	0.2	0.3
	2C′-a	10.8	31.0	8.8	**47.3**	0.0	0.3	0.6	1.2
	2D′-a	9.4	12.8	6.1	**66.1**	0.7	1.0	0.1	3.9

TABLE VI
CONFUSION MATRIX FOR APT2|APT1 RESULT (%)

		Predicted Label				
		MB	**HD**	**HH**	**CD**	**VC**
True Label	**MB**	**78.6**	1.8	15.3	4.2	0.2
	HD	10.2	**68.3**	20.8	0.8	0.0
	HH	9.8	**55.6**	34.4	0.2	0.0
	CD	25.6	0.5	0.0	**72.3**	1.6

TABLE VII
CONFUSION MATRIX FOR APT1|APT2 RESULT (%)

		Predicted Label			
		MB	**HD**	**HH**	**CD**
True Label	**MB**	**79.4**	0.5	18.0	2.2
	HD	0.1	**79.3**	19.6	1.0
	HH	0.0	1.2	**98.6**	0.2
	CD	0.0	0.3	3.9	**95.7**
	VC	0.0	0.3	0.6	**99.1**

VI. SUMMARY

The generalizability of CNN-based supervised learning for inter-floor noise type/position classification in residential buildings was shown. It was shown against two datasets collected in two apartment buildings. The CNN-based model used in this work was VGG16 with an adaptation layer. It was initialized with pre-trained weights for an image recognition task. And it was fine-tuned reducing cross entropy loss and L_2-regularization loss of weights between VGG16 and the adaptation layer. Various tasks were designed and tested to show the generalizability of the method and the viability of type classification knowledge transfer.

ACKNOWLEDGMENT

The authors thank Sangkyum An and Minseuk Park for providing experiment sites.

REFERENCES

[1] C. Maschke and H. Niemann, "Health effects of annoyance induced by neighbour noise," Noise Control Eng. J., vol. 55, pp. 348–356, May 2007.

[2] J. Y. Jeon, "Subjective evaluation of floor impact noise based on the model of ACF/IACF," J. Sound Vib., vol. 241, pp. 147–155, March 2001.

[3] Floor Noise Neighborhood Centre of the Korea Environment Corporation, "2018 Report," available at http://www.noiseinfo.or.kr/about/stats/counselServiceSttus_01.jsp, accessed 29-August-2019.

[4] H. Choi, H. Yang, S. Lee, and W. Seong, "Classification of inter-floor noise type/position via convolutional neural network-based supervised learning," Appl. Sci., vol. 9, pp. 3735, September 2019.

[5] R. Bahroun, O. Michel, F. Frassati, M. Carmona, and J. Lacoume, "New algorithm for footstep localization using seismic sensors in an indoor environment," J. Sound Vib., vol. 333, pp. 1046–1066, February 2014.

[6] J. D. Poston, R. M. Buehrer, and P. A. Tarazaga, "Indoor footstep localization from structural dynamics instrumentation," Mech. Syst. Sig. Processing, vol. 88, pp. 224–239, May 2017.

[7] M. Mirshekari, S. Pan, J. Fagert, E. M. Schooler, P. Zhang, and H. Y. Noh, "Occupant localization using footstep-induced structural vibration," Mech. Syst. Sig. Processing, vol. 112, pp. 77–97, November 2018.

[8] H. Choi, H. Yang, S. Lee, and W. Seong, "Inter-floor noise dataset and code implementation," available at https://github.com/yodacatmeow/indoor-noise, accessed 5-June-2020.

[9] H. Choi, S. Lee, H. Yang, and W. Seong, "Classification of noise between floors in a building using pre-trained deep convolutional neural networks," 16th International Workshop on Acoustic Signal Enhancement (IWAENC), pp. 535–539, September 2018.

[10] M. Abadi *et al*., "TensorFlow: Large-scale machine learning on heterogeneous systems," available at http://tensorflow.org, accessed 13-October-2019.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, April 2014.

[12] O. Russakovsky *et al*., "ImageNet large scale visual recognition challenge," Int. J. Comput. Vision, vol. 115, pp. 211–252, April 2015.

[13] S. J. Pan and Q. Yang, "A Survey on transfer learning," IEEE. T. on. Data Eng., vol. 22, pp. 1345–1359, October 2009.

[14] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1717–1724, June 2014.

[15] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," J. Mach. Learn. Res., vol. 13, pp. 281–305, February 2012.