

Article

# Practical Inter-Floor Noise Sensing System with Localization and Classification

Junho Son <sup>1</sup>, Chong-Min Kyung <sup>2</sup> and Hyuntae Cho <sup>3,\*</sup><sup>1</sup> Samsung Electronics, Hwasung-Si 18448, Korea<sup>2</sup> Department of Electrical Engineering, KAIST, Yuseong-Gu, Daejeon 34141, Korea<sup>3</sup> Center for Integrated Smart Sensors, KAIST, Yuseong-Gu, Daejeon 34141, Korea

\* Correspondence: phd.marine@kaist.ac.kr; Tel.: +82-42-350-8705; Fax: +82-42-350-8700

Received: 12 July 2019; Accepted: 19 August 2019; Published: 21 August 2019



**Abstract:** Inter-floor noise is a severe social problem which causes illegal arson, violence, and even murder. In this paper, an inter-floor noise sensing system is proposed to detect and record information related to inter-floor noise in an apartment building. The proposed system measured the noise level and estimated the direction of the noise source along with the type of noise. The noise level measurement is based on the sound pressure level (SPL) measurement, which is a logarithmic measure of the effective pressure of a sound relative to a reference sound pressure. Noise source localization was performed using the estimated time difference of arrival (TDOA) from the microphone array. For the classification of noise types, the Mel frequency cepstral coefficients (MFCC) and zero-crossing rate (ZCR) were extracted from a noise signal, and the k-nearest neighbor algorithm was used to classify the type of noise. In addition, we developed a noise monitoring hardware to evaluate our methods in the actual environment. The experimental results demonstrated that the proposed system had a reliable accuracy for each functional unit. The results showed that the error of the noise level was approximately  $\pm 1.5$  dB(A), the error of the direction estimation was approximately  $\pm 10^\circ$ , and the accuracy of the classification for the noise type was more than 75%. These output data from the proposed system are expected to be used as important reference data for any dispute cases due to inter-floor noise.

**Keywords:** acoustic noise sensor; inter-floor noise; microphone array; sound classification; sound pressure level; sound source localization

## 1. Introduction

Due to high population density, many countries have a very high rate of apartment housing. In the case of South Korea, the proportion of apartment house among all types of housing has continuously increased and the total percentage of apartment housing reached 75% in 2016 [1]. Roughly speaking, about three quarters of Korean citizens are sharing floors with their neighboring residents in their current housing situation. The unavoidable problem that arises from the high percentage of apartment housing is the inter-floor noise between neighboring residents.

Inter-floor noise in many countries has been a serious social problem. Conflicts between neighboring residents have occurred frequently, and some of them resulted in crimes such as arson, violence, and even murder. The governments of several countries created a legal standard related to inter-floor noise to provide a quantitative definition of it. The legal standard specifies different threshold levels of inter-floor noise according to their type and occurrence time [2,3].

A major problem of current situations is the lack of a specialized device for recording inter-floor noise even though there are a number of noise sensors in the market. With the currently available sound level meters, users can only measure the noise level without localization and classification [4–7].

In addition to the noise level, there are other more important data regarding inter-floor noise which can be used as reference data in dispute cases.

In this paper, an inter-floor noise sensing system, which detects not only the noise level, but also other meaningful information related to inter-floor noise is proposed. The proposed system captures and displays three pieces of information related to inter-floor noise: noise level, three-dimensional direction of the noise source, and type of noise. The noise level measurement is based on a logarithmic measure of the effective pressure of a sound relative to reference sound pressure, and the noise source localization is performed using the estimated time difference of arrival (TDOA) from the microphone array which consists of three pairs. For the classification of noise types, Mel frequency cepstral coefficients (MFCC) and zero-crossing rate (ZCR) were extracted from the noise signal, and k-nearest neighbor algorithm was used to classify the type of noise. Knowing the direction of the noise source helps in finding the location of the point where the inter-floor noise occurs frequently. The classification of the noise type is also important because the legal standard is different according to the type of inter-floor noise that is occurring.

The rest of this paper is organized as follows. Section 2 introduces related works. Section 3 illustrates how each function of the proposed system is implemented with detailed algorithms. Section 4 provides the experimental results of the proposed system. Finally, the conclusion of this paper is presented in Section 5 with recommendations for future studies.

## 2. Related Works

The proposed inter-floor noise sensing system contains three functional blocks: noise level, sour localization, and noise classification. In this chapter, we briefly introduce related works of each functional block of the proposed system.

### 2.1. Noise Level Measurement

Risojevic et al. [8] proposed accurate indoor sound level measurement on a low-power and low-cost wireless sensor node with limited computational resources. They implemented non-calibrated calculation of the sound pressure level and performed digital A-weighting filtering on the node. According to the experimental results, their approach could measure the noise levels of up to 100 dB with a mean difference of less than 2 dB compared to Class 1 sound level meters. Santini et al. [9] developed a system for the collection and logging of noise pollution data based on tiny sensor nodes. Two studies enabled real-time acquisition, processing, and visualization of data collected in wireless sensors networks. Zamora et al. [10] presented a Smartphone-based noise monitoring system. They focused their effort on the sound capture, processing procedure, and analyzing the impact of different noise calculation algorithms. They could measure the noise level of up to 95 dB and reduce average error below 2%. Most previous studies adopted the A-weighting function to improve the accuracy of the system [11,12]. It is defined as a standard in the international standard IEC 61672-1: 2003 [13]. They have complied with the requirement of the international standard in terms of accuracy. However, they did not consider noise source localization and the classification of noise types. These factors can be used as important reference data for any dispute cases due to inter-floor noise.

### 2.2. Sound Source Localization

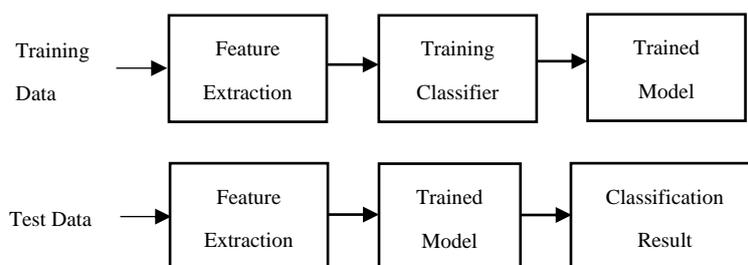
The sound source localization technique is being widely used in various applications such as teleconferencing [14] and robot navigation [15,16]. There have been several investigations that attempted to localize the source of the sound. Among those approaches, the sound source localization based on time difference of arrival (TDOA) from microphone arrays [17–21] is the most widely used and reliable technique. In the proposed sound localization system, the TDOA estimated from the microphone array was used to localize the direction of the noise source. The microphone array structure consisted of several pairs of microphones. Basically, one microphone pair was an angle sensing sensor of the incoming sound signal. To estimate the incident angle of sound, it was assumed that the distance

to the sound source is much larger than the interval of the microphones such that incoming sound waves can be considered as coming in parallel [22–24].

Other studies [25,26] presented localization systems based on the time difference of arrival on Zigbee networks. They achieved high accuracy within several meters even though they use 2.4 GHz radio frequency. These approaches require accurate time synchronization since their localization methods are based on multiple sensor nodes.

### 2.3. Sound Classification

The sound classification technique is also widely used in various fields of applications such as environmental sound recognition [27,28] and speech recognition [29]. Figure 1 provides the basic steps of the classification task. In order to perform the classification, the first requirement is to train a classifier. Meaningful information is extracted from the training data, and this meaningful information is often referred to as a feature. Features are extracted from the raw data because the raw data contain a considerable amount of redundant information and reduction in the dimensionality of the data for later computation is desired. The extracted features were placed onto the feature space, and ‘clustering’ of the labeled data was conducted using the extracted features. This process represents the training of a classifier. The trained model was created as a result of the training process, and the test data was classified in a trained model.

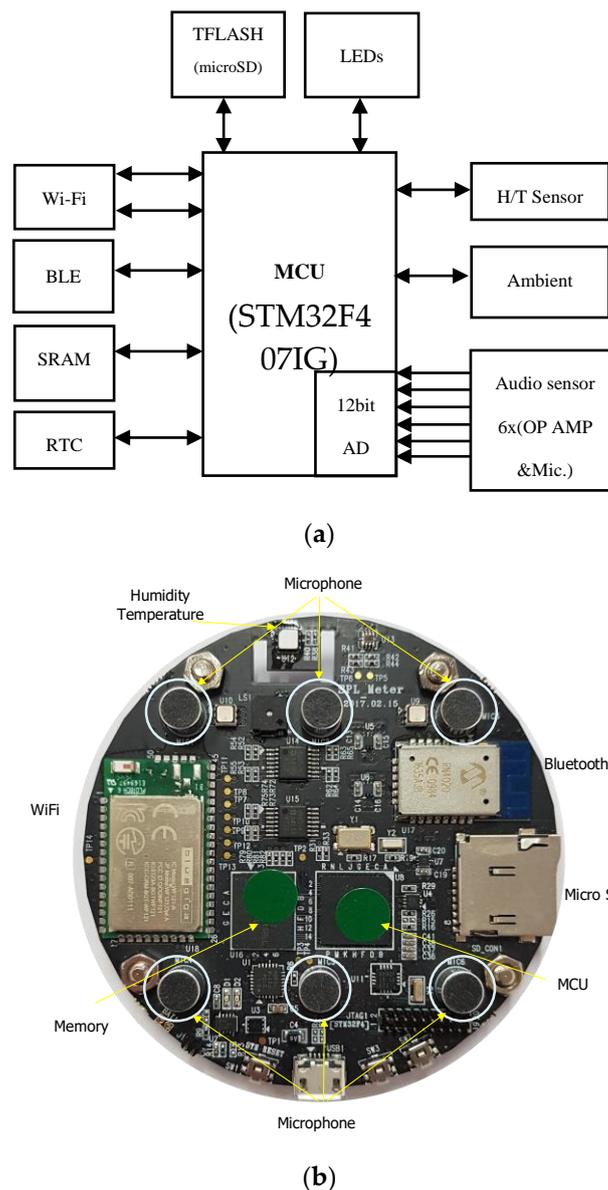


**Figure 1.** Basic methodology of classification task.

After the feature extraction, a classifier should be selected to train a model from the extracted features. There are several classifiers widely used in the field of machine learning such as k-nearest neighbor (k-NN), support vector machine (SVM), convolutional neural network (CNN), recurrent neural network (RNN), and artificial neural network [30–34].

### 3. Implementation of the Inter-Floor Noise Sensing System

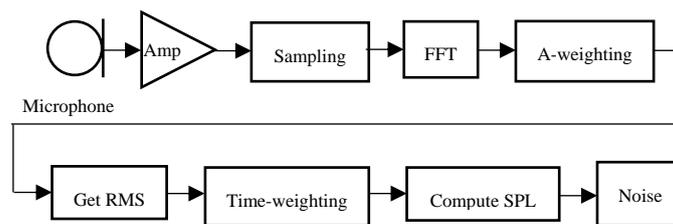
First, a hardware system was designed and implemented to measure the noise by using a microphone array. Figure 2 shows the block diagram and appearance of the system. The system uses a STM32F407IG MCU based on an ARM 32-bit Cortex-M4 MCU that operates at 120 MHz. A 12 bit analog-to-digital converter (ADC) is used to convert analog the audio signal into digital data. Additionally, it contains 2 MB of extra SRAM to improve the computing power and has an external flash memory to store the noise data over the long term. Other components, such as the humidity, temperature sensor, and Bluetooth module are used for our convenience. The inter-floor noise measurement function and sound localization function are implemented in the designed hardware system. The output information (i.e., noise level and direction of the noise) with the raw data is transferred to PC and displayed in real time.



**Figure 2.** Sensor board of the system: (a) block diagram and (b) hardware prototype.

### 3.1. Measurement of Noise Level

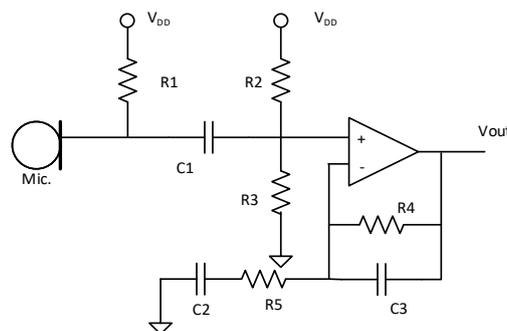
The procedure for estimating the noise level from the noise signal is shown in Figure 3. First, an analog audio signal is converted into an analog voltage signal through a microphone. Sensitivity of the microphone is  $-45$  dB and operating voltage is 2 V. The analog voltage signal is sampled and converted into digital data through the ADC. The fast Fourier transform (FFT) is performed on the digitized samples. In the frequency domain, the A-weighting is applied to consider the human's different perceptions of different frequency levels. After that, the root mean square (RMS) of the signal is calculated. The time weighting is applied to smoothen the displayed data, and finally the sound pressure level (SPL) of the noise is calculated. A detailed explanation of each step will be presented in the following subsections.



**Figure 3.** Flow chart of the sound pressure level (SPL) measurement unit.

### 3.1.1. Amplifying Circuit

Since the output voltage level of the microphone is very low, an amplifying circuit is needed to amplify the voltage produced from the microphones. In the designing process, two things needed to be considered: gain and bandwidth. Gain and bandwidth can be adjusted by changing the values of the resistor and capacitor in the feedback circuit. The audio amplifying circuit in the sensor board is shown in Figure 4. The gain of the amplifying circuit is 20 dB, and the 3 dB cut-off frequency is approximately 11 kHz. The bandwidth of the amplifying circuit should exceed at least 8 kHz, which is a typical frequency range of commercial SPL meters. Therefore, the feedback gain and cut-off frequency should be considered according to the target noise level and bandwidth.



**Figure 4.** Amplifying circuit for the microphone.

### 3.1.2. Sampling and FFT

An amplified voltage signal is sampled in a 12 bit-ADC of the sensor board. The sampling rate is 200 kHz in the proposed system. A much higher sampling rate is used compared to the conventional audio sampling rate (44.1 kHz). This approach was chosen because later in the localization system, a high enough sampling frequency is required to compensate for the very short distance between the microphones (5.7 cm). The Cooley–Tukey algorithm [35], specifically radix-2 FFT is used to calculate the FFT.

### 3.1.3. A-Weighting

A-weighting is applied to consider the human ear's different sensitivity of the different frequencies of sound. The international standard [13] defines the frequency weightings by tables for certain frequencies; but there are appropriate functional realizations of the frequency weightings:

$$R_A(f) = \frac{12194^2 \cdot f^4}{(f^2 + 20.6^2) \sqrt{(f^2 + 107.7^2)(f^2 + 737.9^2)} (f^2 + 12194^2)}, \quad (1)$$

$$A(f) = 20 \log_{10}(R_A(f)) + 2.0. \quad (2)$$

We apply the A-weighting in the frequency domain by calculating the appropriate values of  $A(f)$  for each frequency component.

### 3.1.4. Time-Weighting

After the A-weighting, the RMS voltage of the A-weighted signal is calculated. Then, a time-weighting is applied on the time series of the RMS voltage. A time-weighting is needed to smoothen the output displayed data. The displayed data cannot be read if the rate of data change is too fast. The time weighting is sometimes called exponential averaging, and it acts as low-pass filters to remove the high frequency noise. In the proposed system, the time-weighting of the RMS voltage is expressed as the following equation:

$$V_{RMS} = \alpha V_{RMS\_new} + (1 - \alpha) V_{RMS\_old}, \quad (3)$$

where  $\alpha$  in the above equation is the smoothing factor, and  $0 < \alpha < 1$ . The lower values of  $\alpha$  indicates smoother data displays. The value of  $\alpha$  is given by:

$$\alpha = 1 - \exp\left(\frac{-\Delta T}{\tau}\right), \quad (4)$$

where  $\Delta T$  is the sampling rate of the data and  $\tau$  is time as a constant. Two types of time constants are often used to display the sound pressure level: F (Fast, 125 ms) and S (Slow, 1 s). Both time constants are implemented in the proposed SPL measurement unit.

### 3.1.5. Computing SPL

The final step is to compute the sound pressure level of the signal. It is calculated as:

$$\text{SPL} = 20 \log_{10} \frac{V_{RMS}}{V_{REF}} [\text{dB(A)}], \quad (5)$$

where  $V_{REF}$  is the output RMS voltage when the pressure level of the sound signal is equal to the reference sound pressure  $P_0$  (20 micro-pascal).  $V_{REF}$  is determined from the microphone sensitivity and gain in the amplifying circuit.

## 3.2. Localization of Noise Source

### 3.2.1. Introduction to Time Difference of Arrival

The basic geometry of a microphone pair is shown in Figure 5. Basically, one microphone pair was an angle sensing sensor of the incoming sound signal. To estimate the incident angle of sound, it was assumed that the distance to the sound source is much larger than the interval of the microphones such that incoming sound waves can be considered as coming in parallel. This assumption is reasonable because the interval of the microphones of our system was only 5.7 cm; while, the typical distance to the inter-floor noise source will be at least more than 10 times of the interval of the microphones. The direction angle,  $\theta$ , to the sound source in Figure 5 can be calculated by the following equation:

$$\theta = \cos^{-1}\left(\frac{\tau_{delay} \cdot v_s}{d_{mics}}\right). \quad (6)$$

The TDOA of a microphone pair can be derived from the cross-correlation matrix of samples acquired from each microphone. Generally, the cross-correlation is a measure of similarity of two signals as a function of the displacement of one relative to the other. In the time series analysis, a cross-correlation is often used for estimating the time delay between two signals [24]. The cross-correlation of two sound signals acquired from each microphone  $i$  and  $j$  is expressed as:

$$\text{Corr}_{i,j}[\tau] = \sum_{n=0}^{N-1} x_i[n] x_j[n - \tau], \quad (7)$$

where  $x_i[n]$  is the signal captured by microphone  $i$ . In digital signal processing, the cross-correlation is often calculated in the frequency domain. The major advantage of calculating the cross-correlation in the frequency domain is the reduction in the computation cost due to the use of a fast Fourier transform (FFT). In the frequency domain, cross-correlation is calculated as:

$$Corr_{i,j}[\tau] = FT^{-1}\left[FT^*[x_i] \cdot FT[x_j]\right], \tag{8}$$

where  $FT[\ ]$  and  $FT^{-1}[\ ]$  is the Fourier Transform and the inverse Fourier Transform, respectively.  $FT^*[\ ]$  denotes the complex conjugate of the Fourier Transform. The maximum of the cross-correlation function indicates the point where the signals are best aligned, and the argument of the maximum refers to the delay between the two signals. The delay of arrival between the two signals from microphone  $i$  and  $j$  is derived as:

$$\tau_{delay} = \underset{t}{\operatorname{argmax}}(Corr_{i,j}[t]). \tag{9}$$

We should note that the above result is the delay in a sample if the cross-correlation is computed in a sample domain. To convert the delay in a sample to delay in time, the division of the delay in sample by sampling frequency of the system is required.

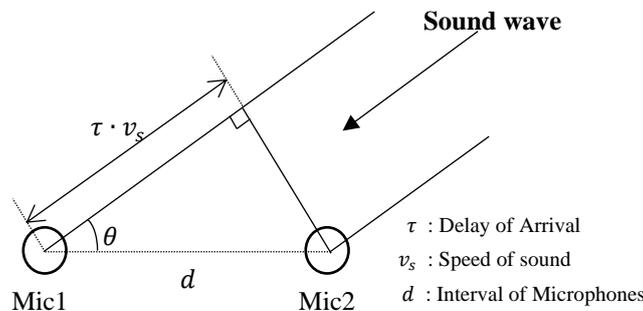


Figure 5. Basic geometry of a microphone pair.

### 3.2.2. Microphone Array Structure

The localization system is roughly divided into two parts. The first part is obtaining the TDOA from the microphone pairs, and the second part is identifying the direction angles from the estimated TDOA. In our localization system, a three-dimensional direction of the sound source is expressed as two angular components: azimuth and elevation. Figure 6 illustrates the two output angles in local Cartesian coordinate of our system.

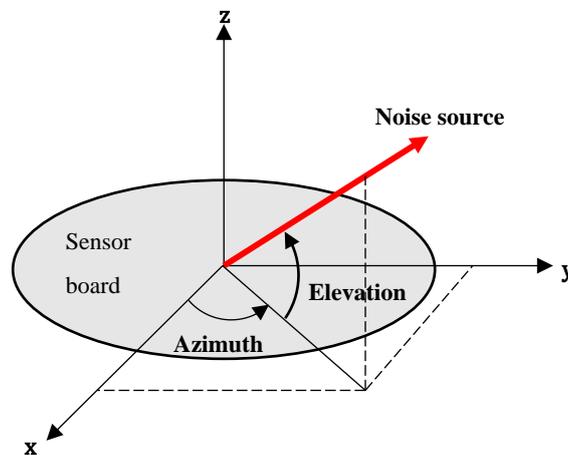
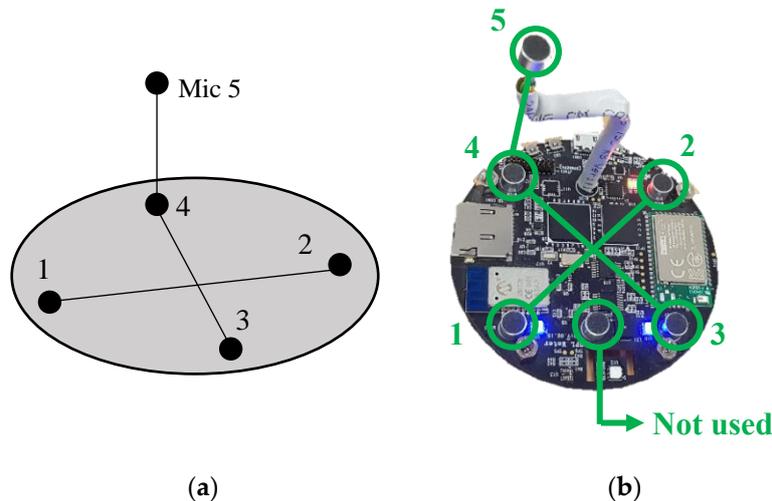


Figure 6. Two output angles of the localization system: Azimuth and elevation.

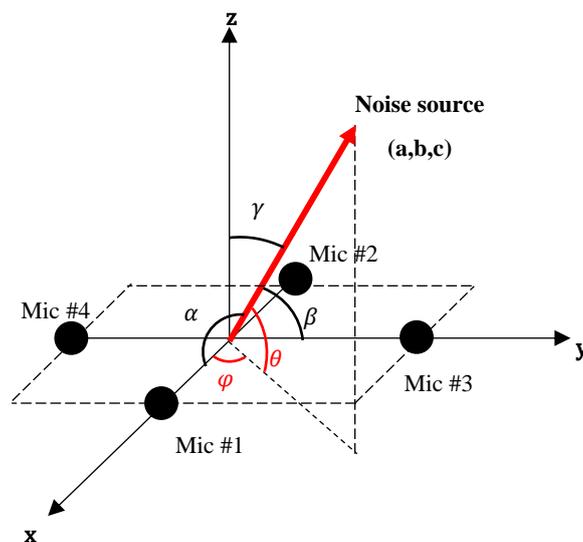
There is a total of six microphones on the hardware system, and only five are used in the proposed microphone array structure. Figure 7 depicts the microphone array structure of our system. A total of five microphones on the sensor board are used to identify the noise source direction in a three-dimensional space. In Figure 7b, the microphones that are connected by a common line formed one microphone pair. Three microphone pairs are used to estimate the TDOAs of the input noise signal. Microphone pair 1 (Mic 1 and 2) and 2 (Mic 3 and 4) are used to estimate the azimuth and elevation. Microphone pair 3 (Mic 4 and 5) is used to remove ambiguity of the  $z$ -axis (decision of  $z > 0$  or  $z < 0$ ) when the elevation is estimated.



**Figure 7.** Pairing of the microphones on the sensor board: (a) simplified microphone array structure and (b) actual hardware system.

### 3.2.3. Estimation of Azimuth and Elevation

The angles of the noise source direction vector with respect to the  $x$ ,  $y$ , and  $z$  axes in a local Cartesian coordinate is exhibited in Figure 8. The  $\alpha$ ,  $\beta$ , and  $\gamma$  are the direction angles of the noise source vector regarding the  $x$ ,  $y$ , and  $z$  axes, respectively. The  $\varphi$  and  $\theta$  denote the azimuth and elevation, respectively. Our goal is to identify the azimuth  $\varphi$  and elevation  $\theta$  from the direction angles  $\alpha$  and  $\beta$ .



**Figure 8.** Direction angles in the local Cartesian coordinate.

From the TDOA estimated from each microphone pair, the value of  $\alpha$  and  $\beta$  are calculated by:

$$\alpha = \cos^{-1}\left(\frac{TDOA_{mic1,2} \cdot v_s}{d_{mics}}\right), \quad (10)$$

$$\alpha = \cos^{-1}\left(\frac{TDOA_{mic1,2} \cdot v_s}{d_{mics}}\right). \quad (11)$$

Following the relationship between the direction cosines of the noise direction vector holds:

$$\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1. \quad (12)$$

The relationship between the elevation  $\theta$  and z-axis direction angle  $\gamma$  is:

$$\theta = 90^\circ - \gamma. \quad (13)$$

We can derive an expression for the elevation from above the two equations:

$$\theta = 90^\circ - \cos^{-1}\left(\sqrt{1 - \cos^2 \alpha - \cos^2 \beta}\right). \quad (14)$$

For azimuth, the tangent of azimuth is expressed as:

$$\tan \varphi = \frac{b}{a} = \frac{\cos \beta}{\cos \alpha}. \quad (15)$$

We can derive an expression for azimuth as:

$$\varphi = \tan^{-1}\left(\frac{\cos \beta}{\cos \alpha}\right). \quad (16)$$

If only two microphone pairs are used to identify the elevation, the ambiguity between the  $z > 0$  plane and  $z < 0$  plane exists. To enable a full 3D space localization, one additional microphone pair along the z-axis is used. The sign of the TDOA of microphone pair 3 determines where the noise source is located, either the  $z > 0$  space or  $z < 0$  space. Considering this issue, the expression of elevation is modified as:

$$\theta = \text{sgn}(TDOA_{mic4,5}) \left[ 90^\circ - \cos^{-1}\left(\sqrt{1 - \cos^2 \alpha - \cos^2 \beta}\right) \right]. \quad (17)$$

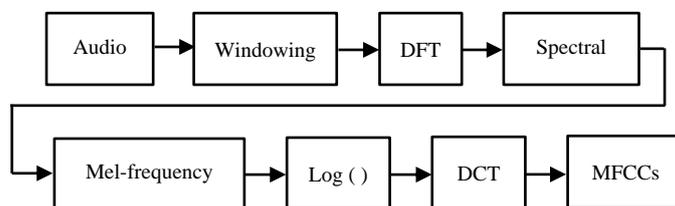
### 3.3. Classification of Noise Types

The procedure for classifying the noise type is basically the same as the procedure explained in Section 2. The first requirement is to extract the meaningful audio features from the audio data. This part is essential because the raw audio data contains a great deal of redundant information. There are several features used in sound classification or recognition. Among those, the most popular and widely used audio feature is the Mel Frequency Cepstral Coefficients (MFCC).

The MFCCs refer the coefficients of the MFC (Mel-frequency cepstrum), and the MFC is a representation of the short-term power spectrum of a sound on the Mel scale. The Mel scale, whose name comes from the word melody, is a perceptual scale of pitches judged by listeners to be equal in distance from one another. As a human's perceptions of pitch intervals are more sensitive in the low frequency region, the Mel scale is a non-linear function in frequency representation. A formula to convert  $f$  (Hz) into the  $m$  (Mel scale) is:

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right) = 1127 \ln\left(1 + \frac{f}{700}\right). \quad (18)$$

Figure 9 displays the flow chart for extracting the MFCCs from audio data. Prior to the MFCC extraction, an audio clip was divided into short-time frames. The typical frame length was usually from 10 to 30 ms with a 50% overlap between the adjacent frames.



**Figure 9.** Flow chart of extracting Mel Frequency Cepstral Coefficients (MFCCs).

After framing, a window function was applied to each frame. A popular choice for a window function is the Hamming window. The Hamming window is defined as:

$$w[n] = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad (19)$$

where  $N$  is the window length. After windowing, a FFT was conducted on each frame. The next step was to apply the triangular filters onto the Mel-scale. Each filter in the filter bank had a triangular form and an amplitude of 1 at the center frequency. The intervals between the center frequencies of each filter was set to be equal in the Mel-scale. The final step was to do a discrete cosine transform (DCT) after taking the logarithm of the filtered spectral magnitudes. As a result, the MFCCs from each filter were obtained. Only the first 12~13 MFCCs were chosen because the higher indexed MFCCs usually did not contain useful information for sound classification.

### 3.3.1. Training Dataset

For training purposes, the public audio dataset named ESC-50 was used [36]. ESC-50 contains 2000 annotated audio clips for environmental sound classification. The dataset is composed of 50 classes of various environmental sounds and each class consists of 40 audio clips. Each audio clip is 5 s long and the sampling rate is 44.1 kHz.

Among the 50 classes, seven classes were chosen that are common types of inter-floor noise in an apartment house. Those classes are (1) footsteps, (2) door knock, (3) dog barking, (4) baby crying, (5) clock alarm, (6) vacuum cleaner, and (7) washing machine. According to the legal standard in Table 1, footsteps and door knock are noises occurring from direct impact, while the rest of the classes are classified as noise delivered through air.

**Table 1.** Classification accuracy of the test set.

Class	Accuracy (%)
Clock alarm	40
Crying baby	100
Dog barking	80
Door knock	80
Footsteps	80
Vacuum cleaner	70
Washing machine	80
Average	75.7

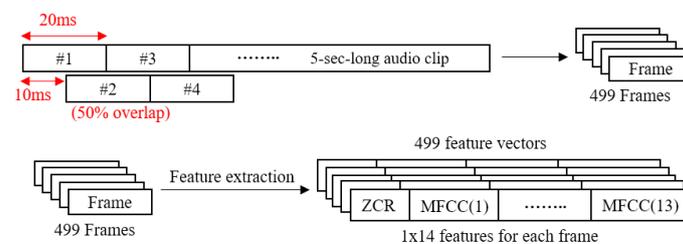
### 3.3.2. Training Process

The feature extraction process of one audio clip is illustrated in Figure 10. First, the original audio clip was divided into 20 ms-long sub-frames with 50% overlap. A total of 499 short frames were obtained from one 5 s-long audio clip.

Generally, the MFCCs feature is a typically used audio classification. However direct impact, such as hammering and children leaping, needed to be detected. Therefore, a combination of the MFCCs and zero-crossing rate is used in our classification system. The zero-crossing rate is the rate of sign-changes along a time signal as its name states. It is known to be a powerful feature in classifying percussive sounds [37]. Just like the MFCCs, audio data is divided into short-time frames prior to feature extraction. For each divided frame, the zero-crossing rate is computed. The zero-crossing rate of the digital audio signal  $x[t]$  is calculated as following equation:

$$\text{ZCR}(x[t]) = \frac{1}{2(N-1)} \sum_{i=1}^{N-1} \{ \text{sgn}(x[i]) - \text{sgn}(x[i-1]) \} \quad (20)$$

These two features are extracted from each frame. For MFCCs, 13 MFCCs were chosen and the remainder were discarded.



**Figure 10.** Feature extraction from one audio clip.

As a result, the  $499 \times 14$  feature table was obtained from one audio clip (size of  $1 \times 220,500$ ). This result is a 97% reduction of the data size. Since the zero-crossing rate and MFCCs are not on the same scale, they needed to be normalized, which was conducted by subtracting the mean and dividing by the standard deviation of each column of the feature table. After the feature extraction, a classifier was trained with the obtained features. A k-NN classifier [38] with  $k = 5$  was used in the proposed system. As its name states, the k-NN algorithm identifies the k nearest neighbor data from the input data in the feature space. The output class was determined as the most common class type among its k nearest neighbor data.

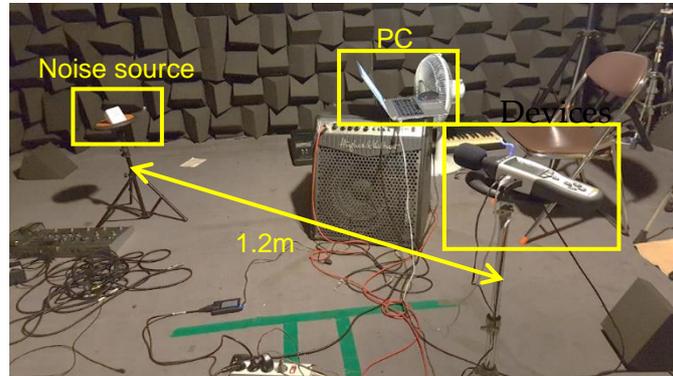
## 4. Experimental Results

### 4.1. Measurement of Noise Level

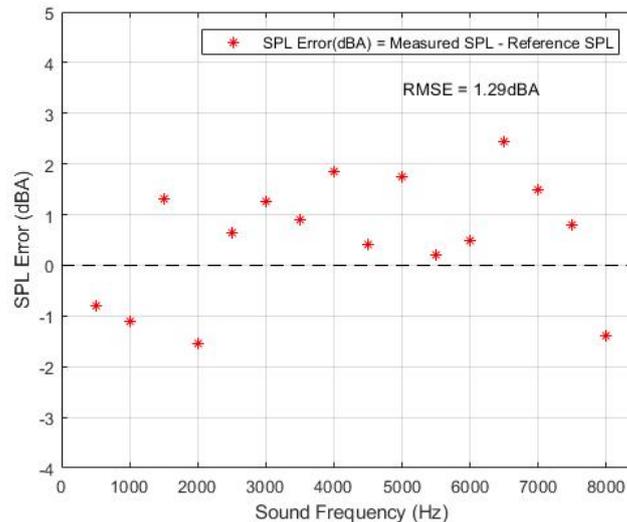
To evaluate the accuracy of the noise level, the SPL value obtained from the proposed system was compared with a commercial SPL meter (DT-8852) [39] as a reference. Figure 11 shows an experimental environment for performance evaluation. The DT-8852 has an error of  $\pm 1.4$  dB(A). Experiments were performed in an anechoic chamber. A Bluetooth speaker was used to generate sound signals, and the distance between the speaker and noise monitoring devices was set to 1.2 m. Pure-tone sinusoidal sound waves with frequencies that ranged from 500 to 8000 Hz were used since the frequency range of commercial SPL meters is usually up to 8000 Hz. The experimental environment in the anechoic chamber and reference SPL meter were used in the experiment.

The measurement error was calculated by subtracting the reference SPL from measured SPL. A total of 16 sound signals with different frequencies were used in the experiment. The frequency intervals of each sound signal was 500 Hz. Figure 12 shows the measurement error of the proposed

system compared to the reference SPL meter. The maximum deviation from the reference value did not exceed 2 dB(A) with the exception of the signal at 6500 Hz. To illustrate the system accuracy in one value, the RMS error from each result was computed. The resulting RMS error in the experiment is 1.29 dB(A).



**Figure 11.** Experimental setup for SPL measurement.

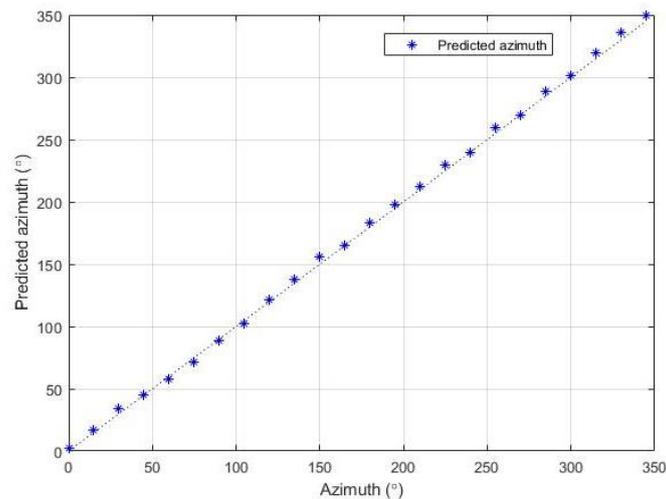


**Figure 12.** Error of the SPL measurement for different frequency levels.

#### 4.2. Localization of Noise Source

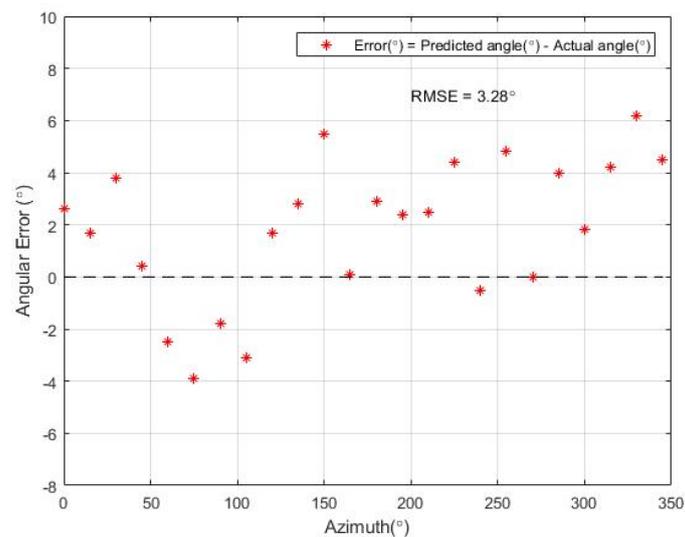
The accuracy of the estimated azimuth and elevation from the proposed system was evaluated in the experiment. The experiment was performed in a reverberant room to provide more realistic conditions of an apartment building environment. A 1 kHz pure-tone sound wave was used as a noise signal. The distance between the speaker and our system in the experiment was set to 50 cm.

In the experiment regarding the estimation of azimuth, the noise source (speaker) was placed at each azimuth from  $0^\circ$  to  $345^\circ$ . Each test angle was  $15^\circ$  apart from each other, so the experiment was performed for 14 different azimuth values. The elevation was set to  $0^\circ$  in the experiment. Figure 13 shows the result of the predicted versus actual azimuth values. The dashed blue line is the ground truth value and blue dots indicate the estimated azimuth values from the system.



**Figure 13.** Predicted azimuth values compared to actual values.

Similar to the SPL error in the previous experiment, the measurement error of the azimuth prediction is calculated as the difference between the predicted and actual azimuth values. Figure 14 illustrates the distribution of the estimation error at each azimuth value.



**Figure 14.** Estimation errors of azimuth for different positions.

The most deviations occurred in the  $\pm 5^\circ$  range from the ground truth. The RMS error of the predicted azimuth value was computed, and the resulting RMS error in the experiment was  $3.28^\circ$ .

In the experiment regarding the estimation of elevation, the noise source (speaker) was placed at each elevation value from  $-90^\circ$  to  $90^\circ$ . Each test angle was  $15^\circ$  apart from each other. Figure 15 shows the predicted azimuth values for each of the actual azimuth values. The dashed line indicates the ground truth and the blue dot is the predicted elevation for each position. Figure 16 illustrates the distribution of the estimation error at each elevation position. The accuracy of the estimated elevation is lower compared to the azimuth estimation. This outcome is due to the system not using the information from the microphone pair along the  $z$ -axis in the estimation process. The angular error is calculated as the difference of the predicted and actual angles. The resulting RMS error of the elevation estimation is  $5.71^\circ$ .

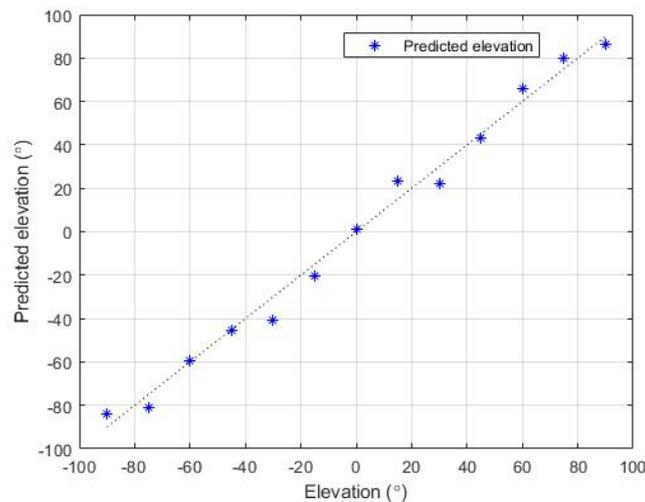


Figure 15. Predicted elevation values compared to actual values.

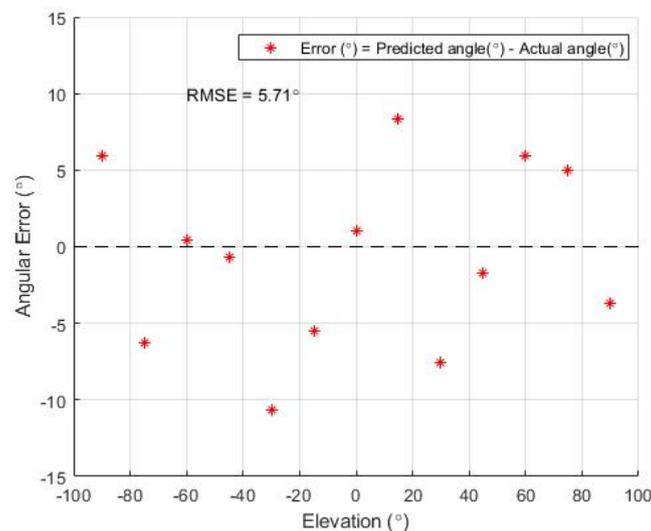


Figure 16. Estimation errors of elevation for different positions.

#### 4.3. Classification of Noise Type

We used seven classes of the ESC-50 dataset for the noise type classification. As earlier mentioned, the selected seven classes were a clock alarm, crying baby, dog barking, door knock, footsteps, washing machine, and vacuum cleaner. They undergo the proposed hardware system to extract exact features. Audio clips in the ESC-50 dataset were used as the training set in our classification system. Two kinds of features (MFCCs and zero-crossing rate) were extracted from each audio clip, of which formed a feature space. The type of inter-floor noise of the test audio clip was determined using k-NN classifier ( $k = 5$ ).

The accuracy of our trained model was evaluated first by performing a five-fold cross-validation. The accuracy obtained by the cross-validation indicates how accurately a trained model performs in test situations. The conceptual diagram of a five-fold cross-validation is shown in Figure 17. In a five-fold cross-validation, the training set is randomly divided into five subsets. Of five subsets, one is used as a test set, while the other subsets are used as training sets. Iteration is repeated until all of the subsets are used once as a test set. Finally, cross-validation accuracy is obtained by averaging the accuracies from all of the iteration processes.

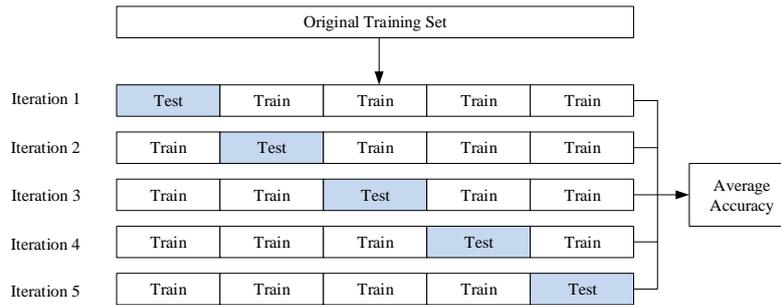


Figure 17. Diagram of five-fold cross-validation.

The classifier was trained in two different ways to see the effect of the zero-crossing rate as a training feature. First the model was trained by using only MFCCs; then a second model was trained by using both the MFCCs and zero-crossing rate. Figure 18 displays the cross-validation accuracy of the trained model (MFCC only). The average cross-validation accuracy of the trained model is 73.6%.

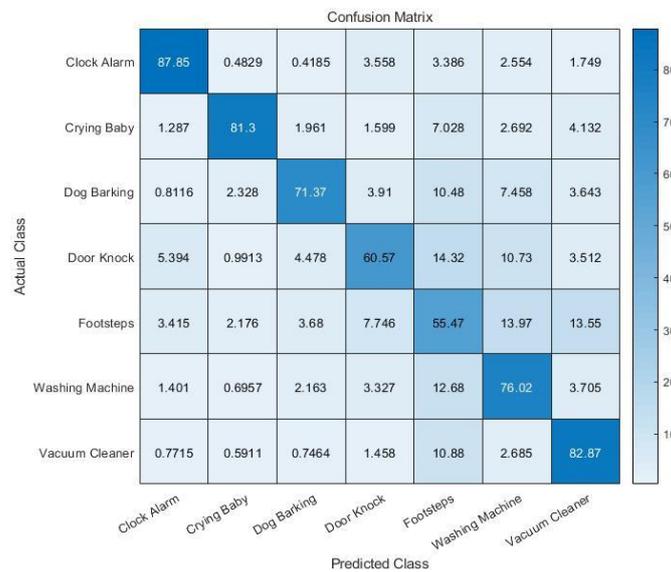


Figure 18. Five-fold cross-validation accuracy (MFCC only).

Figure 19 shows the cross-validation accuracy of the second model (MFCC + zero-crossing rate). Among the seven classes, the sounds from direct impact, such as a door knock and footsteps, had relatively low accuracy. The average accuracy of the five-fold cross-validation is 77.3%, which is about 4% higher than the accuracy of the first model. This result demonstrates that the use of the zero-crossing rate improves the accuracy of the classifier.

A test was conducted after training a classifier. As a test set, the sound samples of the seven classes were primarily collected from YouTube. Then, the actual output sound underwent our hardware system to contain hardware dependent features. For the clock alarm sound, due to the lack of appropriate samples on YouTube, some were manually recorded. The collected sound clips were resized into 5 s long clips using a software named Audacity [40]. As a result, 10 audio clips were prepared for each class as a test set. Figure 20 shows the classification results of the test set. The numbers inside the confusion matrix simply indicates the number of audio clips which belonged to a specific cell of the confusion matrix.

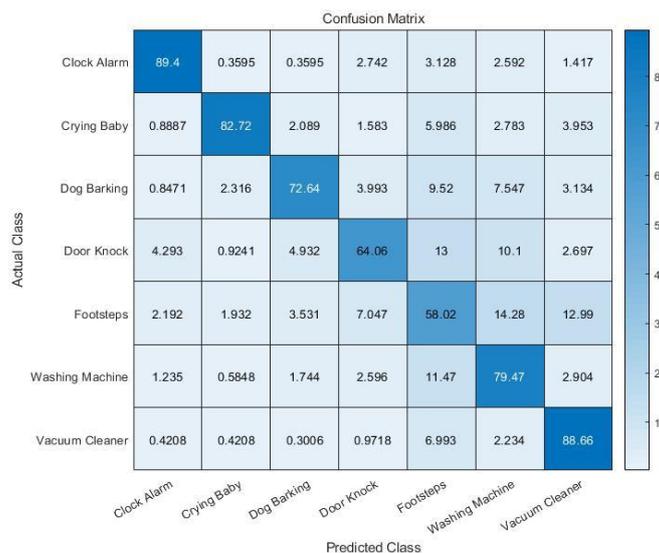


Figure 19. Five-fold cross-validation accuracy (MFCC + zero-crossing rate).

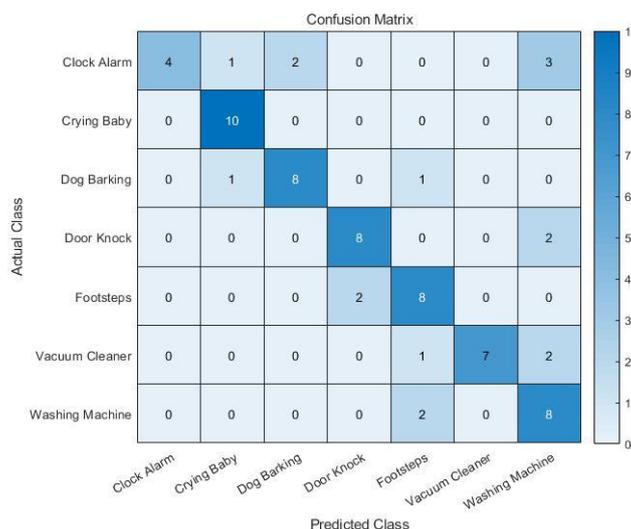


Figure 20. Test results of the trained system.

Among the 10 samples, more than seven samples are classified correctly, with the exception of the clock alarm sound. The classification accuracy of the clock alarm sound is relatively low. This result might be from the lack of variety among the training data for clock alarms. Most of the clock alarm training data just contained a ‘beep’ sound; while the test data of clock alarms contained various kinds of clock alarm signals. The accuracy in the classification indicates the rate of true predictions among all of the samples. The accuracy of each class is given by:

$$\text{Accuracy (\%)} = \frac{\text{Number of correctly predicted samples}}{\text{Total number of samples}} \tag{21}$$

Table 1 shows the classification accuracy of each class on a test set. Except clock alarm, the classes have acceptable accuracy. The average classification accuracy on a test set is 75.7%, which is similar to a five-fold cross-validation accuracy.

A limitation of the current classification system is the lack of training data. Our target classification classes are inter-floor noises; while the training data used in this experiment is for general environmental sound classification. For example, in the footsteps class, the training data contained not only indoor

footsteps, but also footsteps in an outdoor environment, such as a gravelly field. Therefore, the dataset is not optimized for the target purpose of our system. The accuracy of the system can improve further if the training dataset from real inter-floor noise is used.

## 5. Conclusions

In this paper, an inter-floor noise sensing system which captured and displayed important information regarding inter-floor noise was proposed. The proposed system measured the noise level and estimated the direction of the noise source and type of noise. An embedded system with a microphone array was used to capture the noise signal. The noise level measurement and localization system were implemented on the sensor board. The classification system was implemented in a PC due to the lack of computing memory of the sensor board.

Experiments were performed for each functional block to evaluate the accuracy of the proposed system. The SPL measurement unit of the system was demonstrated to have a SPL RMS error of 1.29 dB(A). In the localization system, azimuth and elevation were estimated. The accuracy was evaluated for both angles, and the results showed RMS errors of 3.3° and 5.7° for azimuth and elevation, respectively. In the classification system, a public audio dataset was used as a training set. A five-fold cross-validation of the trained model demonstrated an accuracy of 77.3%, and the test results of the manually collected sound samples had an accuracy of 75.7%. To conclude, the proposed noise sensing system had reliable accuracy for all three functional blocks. The detected information from the proposed system are expected to be used as important reference data for any dispute cases due to inter-floor noise.

Future studies should include further calibration with humidity, temperature, and altitude. We should also improve classification performance by adopting other methods such as RNN, CNN, or deep learning.

**Author Contributions:** Methodology, H.C.; Project administration, C.-M.K.; Writing—original draft, J.S.

**Funding:** This work was supported by the Center for Integrated Smart Sensors funded by the Ministry of Science and ICT as Global Frontier Project (CISS-2011-0031870).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Statistics Korea. Statistics Korea, 2017 Population and Housing Census. Available online: <http://kostat.go.kr/portal/eng/pressReleases/7/2/index.board?bmode=read&bSeq=&aSeq=370994&pageNo=1&rowNum=10&navCount=10&currPg=&searchInfo=&sTarget=title&sTxt=> (accessed on 20 August 2019).
2. Berglund, B.; Lindvall, T.; Schwela, D.H. *WHO Guidelines for Community Noise*; WHO: Geneva, Switzerland, 1999.
3. WHO. *Environmental Noise Guidelines for the European Region*; WHO: Geneva, Switzerland, 2018.
4. KIMO Instruments. DB200 Datasheet. Available online: [https://kimo-instruments.com/sites/kimo/files/2018-10/FTang\\_DB200\\_08-10-18\\_0.pdf](https://kimo-instruments.com/sites/kimo/files/2018-10/FTang_DB200_08-10-18_0.pdf) (accessed on 29 July 2019).
5. TES Electrical Electronic Corp. TES-1352S Datasheet. Available online: [http://www.tes.com.tw/en/product\\_detail.asp?seq=387](http://www.tes.com.tw/en/product_detail.asp?seq=387) (accessed on 29 July 2019).
6. RION. NL-52/51 Datasheet. Available online: [https://rion-sv.com/products/NL-52\\_42-E.html](https://rion-sv.com/products/NL-52_42-E.html) (accessed on 29 July 2019).
7. Noiseaware. Noiseaware Indoor Sensor. Available online: <https://noiseaware.io/> (accessed on 29 July 2019).
8. Risojevic, V.; Rozman, R.; Pilipovic, R.; Censnovar, R. Accurate indoor Sound Level Measurement on a Low-Power and Low-Cost Wireless Sensor Node. *Sensors* **2018**, *18*, 2351. [CrossRef] [PubMed]
9. Santini, S.; Ostermaier, B.; Vitaletti, A. First Experiences Using Wireless Sensor Networks for Noise Pollution Monitoring. In Proceedings of the Workshop on Real-World Wireless Sensor Networks, Glasgow, Scotland, 1 April 2008; pp. 61–65.

10. Zamora, W.; Calafate, C.T.; Cano, J.-C.; Manzoni, P. Accurate Ambient Noise Assessment Using Smartphone. *Sensors* **2017**, *17*, 917. [[CrossRef](#)] [[PubMed](#)]
11. Nilsson, M.E. A-weighted sound pressure level as an indicator of short-term loudness or annoyance of road-traffic sound. *J. Sound Vib.* **2007**, *302*, 197–207. [[CrossRef](#)]
12. Bech, S.; Zacharov, N. Appendix D: A-, B-, C-and D-Weighting Curves. In *Perceptual Audio Evaluation-Theory, Method and Application*; John Wiley & Sons: Hoboken, NJ, USA, 2006; pp. 377–437.
13. IEC. IEC 61672-1: 2013 Electroacoustics-Sound Level Meters-Part 1: Specifications. Available online: [https://infostore.saiglobal.com/preview/98701543932.pdf?sku=856357\\_SAIG\\_NSAI\\_NSAI\\_2037190](https://infostore.saiglobal.com/preview/98701543932.pdf?sku=856357_SAIG_NSAI_NSAI_2037190) (accessed on 20 August 2019).
14. Addeo, E.J.; John, D.R.; Gennady, S. Sound Localization System for Teleconferencing Using Self-Steering Microphone Arrays. U.S. Patent 5335011, 2 August 1994.
15. Mumolo, E.; Nolich, M.; Vercelli, V. Algorithms for acoustic localization based on microphone array in service robotics. *Robot. Auton. Syst.* **2003**, *42*, 69–88. [[CrossRef](#)]
16. Valin, J.M.; Michaud, F.; Rouat, J.; Letourneau, D. Robust sound source localization using a microphone array on a mobile robot. In Proceedings of the International Conference on Intelligent Robots and Systems, Las Vegas, NV, USA, 27–31 October 2003; pp. 1228–1233.
17. Chen, X.; Chen, Y.; Cao, S.; Zhang, L.; Chen, X. Acoustic Indoor Localization System Integrating TDMA + FDMA Transmission Scheme and Positioning Correction Technique. *Sensors* **2019**, *19*, 2353. [[CrossRef](#)]
18. Omologo, M.; Svaizer, P. Acoustic event localization using a crosspower-spectrum phase based technique. In Proceedings of the International Conference Acoustics, Speech, and Signal Processing (ICASSP-94), Adelaide, Australia, 19–22 April 1994; Volume 2, pp. 273–276.
19. Omologo, M.; Svaizer, P. Acoustic source location in noisy and reverberant environment using CSP analysis. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP-96), Atlanta, GA, USA, 9 May 1996; Volume 2, pp. 921–924.
20. Svaizer, P.; Matassoni, M.; Omologo, M. Acoustic source location in a three-dimensional space using crosspower spectrum phase. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP-97, Munich, Germany, 21–24 April 1997; Volume 1, pp. 231–234.
21. Brandstein, M.S.; Adcock, J.E.; Silverman, H.F. A closed-form location estimator for use with room environment microphone arrays. *IEEE Trans. Speech Audio Process.* **1997**, *5*, 45–50. [[CrossRef](#)]
22. Handzel, A.A.; Krishnaprasad, P.S. Biomimetic Sound-Source Localization. *IEEE Sens. J.* **2002**, *2*, 607–616. [[CrossRef](#)]
23. Khyam, M.O.; Ge, S.S.; Li, X.; Pickering, M.R. Highly Accurate Time-of-Flight Measurement Technique Based on Phase-Correlation for Ultrasonic Ranging. *IEEE Sens. J.* **2017**, *17*, 434–443. [[CrossRef](#)]
24. Chen, L.; Liu, Y.; Kong, F.; He, N. Acoustic source localization based on generalized cross-correlation time-delay estimation. *Procedia Eng.* **2011**, *15*, 4912–4919. [[CrossRef](#)]
25. Cho, H.; Kim, J.; Baek, J. Practical localization system for consumer devices using Zigbee networks. *IEEE Trans. Consum. Electron.* **2010**, *56*, 1562–1569. [[CrossRef](#)]
26. Bahl, P.; Padmanabhan, V.N. RADAR: An inbuilding RF-based user location and tracking system. In Proceedings of the INFOCOM 2000, Tel Aviv, Israel, 26–30 March 2000; pp. 775–784.
27. Chu, S.; Narayanan, S.; Kuo, C.-C.J. Environmental sound recognition with time–frequency audio features. *IEEE Trans. Audio Speech Lang. Process.* **2009**, *17*, 1142–1158. [[CrossRef](#)]
28. Cowling, M.; Sitte, R. Comparison of techniques for environmental sound recognition. *Pattern Recognit. Lett.* **2003**, *24*, 2895–2907. [[CrossRef](#)]
29. Muda, L.; Begam, M.; Elamvazuthi, I. Voice Recognition Algorithms Using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. *arXiv* **2010**, arXiv:1003.4083.
30. Garcia, V.; Debreuve, E.; Barlaud, M. Fast k nearest neighbor search using GPU. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW'08, Anchorage, AK, USA, 23–28 June 2008; pp. 1–6.
31. Larose, D.T.; Larose, C.D. k-nearest neighbor algorithm. In *Discovering Knowledge in Data: An Introduction to Data Mining*; John Wiley & Sons: Hoboken, NJ, USA, 2014.
32. Keller, J.M.; Gray, M.R.; Givens, J.A. A fuzzy k-nearest neighbor algorithm. *IEEE Trans. Syst. Man Cybern.* **1985**, *4*, 580–585. [[CrossRef](#)]

33. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Neural Information Processing Systems Conference, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
34. Mikolov, T.; Karafiat, M.; Burget, L.; Cemocky, J.; Khudanpur, S. Recurrent Neural Network Based Language Model. In Proceedings of the International Speech Communication Association, INTERSPEECH 2010, Chiba, Japan, 25–30 September 2010; pp. 1045–1048.
35. Cooley, J.W.; Tukey, J.W. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* **1965**, *19*, 297–301. [[CrossRef](#)]
36. Piczak, K.J. ESC: Dataset for environmental sound classification. In Proceedings of the 23rd ACM international conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 1015–1018.
37. Gouyon, F.; Pachet, F.; Delerue, O. On the use of zero-crossing rate for an application of classification of percussive sounds. In Proceedings of the COST G-6 conference on Digital Audio Effects, DAFX-00, Verona, Italy, 7–9 December 2000; p. 26.
38. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]
39. CEM Instruments. DT8852 Datasheet. Available online: <http://www.cem-instruments.com/en/> (accessed on 7 July 2019).
40. Audacity. Available online: <https://www.audacityteam.org/Audacityorg> (accessed on 7 July 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).