Acoustic-Based Position Estimation of an Object and a Person Using Active Localization and Sound Field Analysis

Kihyun Kim 1,2, Semyung Wang 1,* , Homin Ryu 2 and Sung Q. Lee 3

1 School of Mechanical Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Korea; kihyunkim@gist.ac.kr or khyun.kim@lge.com
2 Chief Technology Officer, LG Electronics, Seoul 06763, Korea; homin.ryu2@gmail.com
3 Intelligent Sensors Research Section, Electronics Telecommunication Research Institute (ETRI), Daejeon 34129, Korea; hermann@etri.re.kr
* Correspondence: smwang@gist.ac.kr; Tel.: +82-62-715-2390

Received: 31 October 2020; Accepted: 16 December 2020; Published: 18 December 2020

Abstract: This paper proposes a new method to estimate the position of an object and a silent person with a home security system using a loudspeaker and an array of microphones. The conventional acoustic-based security systems have been developed to detect intruders and estimate the direction of intruders who generate noise. However, there is a need for a method to estimate the distance and angular position of a silent intruder for interoperation with the conventional security sensors, thus overcoming the disadvantage of acoustic-based home security systems, which operate only when sound is generated. Therefore, an active localization method is proposed to estimate the direction and distance of a silent person by actively detecting the sound field variation measured by the microphone array after playing the sound source in the control zone. To implement the idea of the proposed method, two main aspects were studied. Firstly, a signal processing method that estimates the position of a person by the reflected sound, and secondly, the environment in which the proposed method can be operated through a finite-difference time-domain (FDTD) simulation and the acoustic parameters of early decay time (EDT) and reverberation time (RT20). Consequently, we verified that with the proposed method it is possible to estimate the position of a polyvinyl chloride (PVC) pipe and a person by using their reflection in a classroom.

Keywords: active localization; acoustic-based security system; steered response power; sound field analysis; finite-difference time-domain

1. Introduction

With the rapid development of smart homes and voice-assistant technologies, home environments have been established in which loudspeakers and microphones are deployed as sensors or are built-in and distributed through home appliances. The aim of this research was to develop an acoustic-based home security system in the aforementioned environment. An example is shown in Figure 1.

Smart home technology has been evolving to provide proactive services through the monitoring of residents. Therefore, accurately recognizing the scenario in a home environment through a combination of various sensors is important. In [1], studies on context awareness for indoor activity recognition using binary sensors, cameras, radio-frequency identification, and air pressure sensors were reviewed.

A study proposed to recognize each living activity of a user by combining the power meters of appliances with an ultrasonic sensor [2]. In [3], a study was conducted to recognize the complex activities of a kitchen using one module with various sensors.
In such a smart home environment, microphones are used for context awareness and health monitoring owing to their advantage of operating with low power [4]. Dahmen et al. explained that a microphone can be used to identify the scenario of a home environment based on unusual loud noise and the sound of a human falling [5]. In addition, a study explored the possibility of personal identification through footsteps [6].

Automated home security systems have been developed using smart home technologies. Recent home security systems protect residents and their properties, making them safe from intruders as conventional security systems, and they enable the detection of risks to the residents in advance through context awareness of the home environment [5,7].

Microphones in a home security system are primarily used for two purposes: event detection and the classification of unusual sounds, and intrusion detection.

In [8], related studies were reviewed through a comprehensive survey of background surveillance, event classification, object tracking, and situation analysis, and the detection of events in a highly noisy environment was proposed [9]. In [10–12], a microphone array and security camera were combined to detect the sound from an intruder and tilt the security camera in the direction of the sound. Research has been conducted to predict the state of a control space by recognizing the type of sound, analyzing and classifying the sound, and estimating the angular position of the unusual sound using a microphone array [13,14]. A method to identify human behavior in a control space by applying a microphone array to a sound-steered camera was proposed [15].

Intrusion detection using microphones is as effective as the use of security cameras in terms of detecting moving objects [4], and the related studies are summarized below. Studies on intrusion detection have been conducted to determine an intrusion in a security zone based on the change in the room transfer function [16], the sound field variation according to the acoustical transmission path of distributed microphones [17], and the coherence responses in low-frequency environments [18].

However, the conventional methods for event detection have the disadvantage of operating only when a loud noise is generated because the position is determined in the direction of the generated sound, and current techniques for intrusion detection have the disadvantage of only detecting intrusion but not providing the location.

To overcome these disadvantages, we propose an acoustic-based active localization and analysis method to estimate a silent intruder. This study provides a link between localization and intrusion detection techniques using an acoustic-based security system. The reason is that if a person’s position can be estimated and tracked using microphones and loudspeakers, the entry of an unauthorized person into the security space can be known. However, this study primarily addressed the estimation of the position of a silent intruder.
The process of a home security system can be divided into sensing, assessing, and responding. Sensing is very important because it functions as a trigger to operate the security system. Thus, the sensors must be interoperable with each other [5], with a combination of various individual sensors [5,7], or with the information measured by one sensor module [19].

Therefore, through this study, we expected to increase the utilization of microphones used in home security systems. This is because the data measured by a conventional linear microphone array provide only angular information. However, the proposed method also provides the distance, which increases the number of scenarios that can be combined with the information of other sensors.

We present two examples of complementary sensing. In the first one, passive infrared (PIR) sensors function as triggers to awake the security system and record the intrusion using a camera [20]. However, PIR sensors have the disadvantage of being unable to detect an intruder who does not move, moves slowly, or uses heat-insulating clothes. IR sensors have limitations that often cause errors because of their nonlinear sensitivity and the effects of nearby objects [21]. Therefore, if the acoustic-based intrusion detection in [16–18] is applied to the security system to compensate for the weakness in IR sensors, the two sensing systems can complement each other to increase the robustness of the intrusion detection. In the second example, when a microphone array detects the direction of an event, a pan-tilt-zoom (PTZ) camera is rotated and focused on the region of interest [8]. However, because the camera has misrecognitions owing to poor resolutions, distant targets, changes in illuminations, or occlusions [22], the PTZ camera can be operated robustly by providing a distance and an angle of the intruder based on the proposed method.

Therefore, to overcome the shortcomings of conventional acoustic-based intrusion detection systems and achieve the complementary intrusion detection system proposed in [5], this paper describes our proposed active localization method that estimates the position (distance and angle) of a silent intruder using a generated reflection. The main concept is that a loudspeaker generates a signal in the security space. The microphone array extracts the changed signals owing to the intruder, and then the distance and direction are estimated using the changed signals (sound field variation).

Echolocation is a technology that detects a location through an echo which is emitted from a sound source and then returns, and it has been primarily implemented using an ultrasonic sensor. In [23], a biomimetic sonar system was mounted on a robot arm to recognize an object through the vector of the echo envelope. A biomimetic study was conducted to estimate the distance and angle [24]. The distance was estimated using the time delay between the maximum activity owing to the call and the activity owing to the echo, and the angle was predicted by comparing the directivity pattern of the sensor using the notch pattern in the frequency range. Ultrasonic sensors are acoustic sensors used in conventional home security systems. Ultrasonic sensors are active sensors that send signals in a straight line; therefore, the source and receiver can be placed face-to-face [21] or in the same direction to physically detect the intruder [25]. However, owing to the straightness of the signal, they have the disadvantages of utilizing several sensors to increase the detection rate [26] and being unable to detect a person that passes behind an obstacle.

The proposed active localization in the audible frequency utilizes the phenomenon of scattering rather than straightness. Through fundamental research, we verified that the scattering phenomenon in the audible frequency can be used to detect an object [27] or a person hiding behind an obstacle (the related results are described in Appendix B).

We expect that the combination of ultrasound with its straightness and audible sound with strong scattering can detect a person better. Thus, to create a function as a sensor using a loudspeaker and a microphone array, we studied which room conditions result in the reflection generated by an intruder being considered as a new sound source.

We introduce two main topics to implement the proposed idea. The first aspect is signal processing to estimate the position using reflection, and the second involves the simulation and analysis method of the sound field to estimate the position through the reflected sound in the reverberation space. Thus, analysis equations using acoustic parameters are proposed.
When estimating the position of a person using an active acoustic-based method, the analysis of the sound field to determine the position of the intruder has the following implications. In a reverberant environment, the proposed method is not aimed at estimating the position by increasing the number of microphones. In other words, this does not mean that many microphones are distributed in the control space or that the microphone arrays are arranged at each corner of the control space. By using limited hardware, one loudspeaker and one microphone array, the method of estimating a person’s position using the reflected sound is possible through sound field analysis. Therefore, the active localization method proposed in this paper was verified by estimating the position of a polyvinyl chloride (PVC) pipe and a person in a classroom using signal processing and sound field analysis.

The remainder of this paper is organized as follows. In Section 2, the signal model for position estimation using the reflection sound is presented; subsequently, the algorithm is proposed. The feasibility results of the proposed method are presented through the testing of an anechoic chamber. In Section 3, the simulation results for a reverberant environment are described, and the operating conditions in the reverberant space are proposed based on acoustic parameters. In Section 4, the examination of the proposed method using a PVC pipe and a person in a classroom is described. Finally, the conclusions are presented in Section 5.

2. Implementation of Active Localization: Signal Model, Processing, and Feasibility Test

2.1. Signal Model and Definition of Sound Field Variation

The implementation of active localization to estimate the position of a silent intruder requires a reflected sound generated by a silent intruder. We define the sound field variation as the difference between the sound field before intrusion and the sound field after intrusion.

Therefore, the proposed active localization based on the sound field variation can be tested using two steps. The first step is to measure the sound field in a targeted security space using an active approach with a loudspeaker and a microphone array. The second step is to obtain the position of the silent intruder by acquiring the signals of the sound field variation based on a comparison between the signal of the sound field before intrusion (the reference sound field) and after intrusion (the event sound field).

Figure 2 shows the scheme of sound field variation and, as an example, shows some of the reflections. Because the proposed active localization method uses the time signals from a direct sound to the early reflections and we assume that the silent intruder affects the specific reflection locally, we define the decomposition of room impulse responses as in Equations (1) and (2).

![Figure 2](image_url)

Figure 2. Scheme of sound field variation between (a) a reference scenario and (b) an event scenario: (a) Early reflection of room impulse response in a reference scenario without an intruder; (b) early reflection of room impulse response in an event scenario with an intruder.

Equations (1) and (2) represent the decomposition of the room impulse response (RIR) of the reference and intruder scenarios in the time domain, respectively.

\[ h_{\text{ref}} = h_s + h_{r1} + \cdots + h_{rn} + h_{\text{reverberation}} \] (1)
\( h_{\text{event}} = h_{\text{ref}} + \alpha_1 h_{\text{r}_1} + \cdots + \alpha_n h_{\text{r}_n} + h_{\text{person}} + h_{\text{reverberation}} \)  
\hspace{1cm} (2)

where \( h_{\text{ref}} \) is the RIR of a reference scenario, \( h_{\text{event}} \) is the RIR of an event scenario, \( h_{\text{r}_n} \) represents the early reflections of each scenario, \( h_{\text{person}} \) is the new response generated by a person, \( h_{\text{reverberation}} \) is the late reverberation of the room impulse response, and \( \alpha_n \) represents the attenuation coefficients.

Methods to estimate the room shape or locate a sound source by analyzing the echo components of the RIR have been proposed [28–30]. However, because these methods are performed assuming that the RIR is known, the problem of measuring RIR every time an intruder moves in a scenario exists, and they have the disadvantage of being slow systems.

Therefore, in this study, the signal modeling was represented by the viewpoint of the echo decomposition of the RIR, but the signal generated by the loudspeaker was determined using the Gaussian-modulated sinusoidal pulse in Equation (10) to analyze the changed sound field before and after the intrusion, and the extraction of the changed echo component was performed using Equation (3).

If the silent intruder affects the reflection \( h_{\text{r}_n} \) of the RIR locally, the sum of early reflections in an event scenario is approximate to the sum of early reflections in a reference scenario, i.e., \( \alpha_1 \approx \alpha_2 \approx \cdots \approx \alpha_n \approx 1 \). Therefore, Equation (2) can be rewritten as Equation (3).

\[ h_{\text{event}} = h_{\text{ref}} + h_{\text{person}} + \text{error} \]  
\hspace{1cm} (3)

The sound field variation can be calculated using Equation (4).

\[ \Delta H_m = H_{\text{event}}^m - H_{\text{ref}}^m = \frac{G_m}{X} - \frac{Y_m}{X} = \frac{(G_m - Y_m)}{X} = \frac{R_{\text{effect}}^m}{X} \]  
\hspace{1cm} (4)

where \( H_{\text{ref}}^m \) is the transfer function of the control area under the reference scenario shown in Figure 2a, \( H_{\text{event}}^m \) is the transfer function under the event scenario shown in Figure 2b, \( X \) is the input signal, \( G_m \) represents the signals measured by the microphone array after an intrusion, \( Y_m \) represents the reference signals before an intrusion, \( R_{\text{effect}}^m \) represents the changed spatial effects, and \( m \) is a microphone index.

The spatial effects \( R_{\text{effect}}^m \) are assumed to include the sound signals emitted as reflections by the silent intruder. In other words, \( R_{\text{effect}}^m \) can be assumed to also consider the new sound source. This is because the intruder changes the sound field formed from the sound source of a loudspeaker, and then the intruder’s position is estimated using the measured \( R_{\text{effect}}^m \) at a microphone array. This is the same concept in which the incident, reflected, and transmitted phenomena of pressure distribution on the flat surface of a discontinuity are considered to be the sum of the blocked pressure and the radiation pressure in [31]. If the blocked pressure is the signal of the reference scenario in the control space and the radiation pressure is the signal of the event scenario, we can consider it as a new sound source because only the radiation signal remains when the reference scenario signal is removed from the measured signal. From this concept, the loudspeaker is the sound source that generates the sound field in a control area, whereas in the proposed approach, the sound wave formed by the intruder is a new source and the location of the silent intruder can be detected.

2.2. Proposed Algorithm Based on Steered Response Power with Moving Average

In this section, the approach of an algorithm using the steered response power (SRP) is addressed. SRP is a sound source localization technique, and it is known as a robust localization technique in reverberant environments [32,33].

\[ P_k(\theta) \equiv \int_{kT}^{(k+1)T} \left| \sum_{m=1}^{M} w_m \hat{s}_m(t-\hat{\tau}_m(\theta)) \right|^2 dt \]  
\hspace{1cm} (5)
The estimated distance of an intruder is calculated using Equation (9) through the time information corresponding to the peak of the sound field variation.

$$\hat{\theta}_{s} = \arg\max_{\theta} P_k(\theta)$$

where $P_k(\theta)$ is the power value of the classical SRP, $\theta$ is the steered angle, $\hat{\theta}_{s}$ is the look direction, $s_m$ is the microphone signal, $\hat{\tau}_m$ is the delay of each microphone, $W_m$ is the weight, $M$ is the number of microphones, $m$ is the microphone index, $k$ is the block index, and $T$ is the length of some finite-length block signals.

Equations (5) and (6) are the classical SRP using a microphone array, Equation (5) indicates the integrated output of the steered beamformer, and Equation (6) indicates the direction of the sound source.

The proposed active localization estimates the position of a silent person as an angle and a distance in the horizontal plane of a linear microphone array (Figure 3). In other words, the proposed algorithm should represent a two-dimensional plane. In [34,35], the generalized cross-correlation–phase transform (GCC–PHAT) was used to represent the spatial energy map. However, since the PHAT method revealed that the sound source can be determined well under low noise [36], the localization performance in the two dimensions is not robust. The proposed algorithm uses the reflection to estimate the position; thus, the signal-to-noise ratio (SNR) is not high. Therefore, the energy map is expressed by applying the delay and sum beamformer to the classical SRP and a moving average to the power of the steered block signal. Accordingly, Equations (5) and (6) are modified as Equations (7) and (9) to represent the energy map on the horizontal plane of the linear microphone array.

$$P(t, \theta_d) = \left( \frac{1}{N_L} \sum_{n_L=0}^{N_L-1} \sum_{m=1}^{M} | \sum_{t=1}^{T} \hat{\theta}_m(\theta_d) |^2 \right)$$

$$P(t, \theta_d) = \arg\max_{\theta_d} P(t, \theta_d)$$

$$\hat{r}_{s} = \frac{(t_{s} - t_{ref}) \times c}{2 \times f_s}$$

where $P(t, \theta_d)$ is the energy map of the SRP, $\theta_d$ is the set of desired angles, $N_L$ is the length of the moving average, $P(t, \hat{\theta}_s)$ denotes the position results, $\hat{\theta}_s$ is the index of the reflected time sample, $\hat{r}_{s}$ is the estimated distance between the maximum point and the origin, $t_{ref}$ is the index of the peak point generated signal (the origin), $\hat{\theta}_s$ is the estimated angle, $c$ is the speed of sound, and $f_s$ is the sampling frequency.

Figure 3. Example of an application of the proposed active localization method to a room.
In this study, the input value of the SRP used the changed signal between the reference signal and the measured signal. In other words, the impulse response in Equation (3) was not directly predicted, but the sound field variation in the same reproduction signal was estimated by subtracting the reference signal from the measured microphone signal. We used a triangular moving average of 36 samples in the 48 kHz sampling rate, and the estimated distance was calculated as the product of time and sound speed. This averaging method empirically reduced the error variance of the estimated angle and distance in the proposed active localization.

Figure 5 shows the block diagram used to implement the proposed method using the sound field variation and the SRP with a moving average. Figure 5a shows the steps to synchronize the measured signals. Figure 5b shows that the measured signals are stored as the reference signals if no event is detected, as depicted in Figure 5c, and Figure 5d indicates the proposed SRP to estimate the position of a silent person.

In the signal synchronization step in Figure 5a, we set up the block diagram to minimize the time delay between the reference signal and event signal for each microphone. Thus, two steps were involved. The first was to reduce the quantization error by setting the clocks of the loudspeaker and microphone board identically in hardware. The second step, after measurement, was to verify and compensate for the time delay between \( t_{\text{ref}} \) of Equation (7) and the peak of the generated signal based on correlation. The event detection in Figure 5c was used to determine intrusion by selecting the threshold of sound field variation in [17]. In this study, we focused on the analysis of the SRP.
results in Figure 5d. In other words, we aimed to analyze the relationship between the variables (reverberation time and early decay time) in the control space and the signal processing results.

The signal generated by the loudspeaker formed a sound field with a specific frequency band in a security area using the Gaussian-modulated sinusoidal pulse of Equation (10), and then the change to the sound field was measured using the microphone array.

\[
x(t) = Ae^{-\kappa(t-d)^2} \cos(2\pi f_{\text{center}}(t - d))
\]

where \(A\) is the magnitude of the signal, \(\kappa = 5\pi^2 b^2 f_{\text{center}}^2 / (q \ln(10))\) is the envelope constant, \(b\) is the normalized bandwidth, \(q\) is the attenuation of the signal, \(f_{\text{center}}\) is the center frequency, and \(d\) is the time delay.

In this study, the center frequency was fixed at 1 kHz, and the attenuation and normalized bandwidth of the sound source were set to 6 and 0.25, respectively.

The center frequency was 1 kHz because the directivity pattern of the loudspeaker used in the experiment was cardioid at 1 kHz.

When analyzing a short-period pure-tone signal as a frequency component, a discrete-time Fourier transform was used, and at least five periods were required to estimate the frequency components. Therefore, the attenuation and normalized bandwidth were selected to form five periods in the pulse sound (Figure 6).

![Gaussian-modulated sinusoidal pulse](image.png)

**Figure 6.** Gaussian-modulated sinusoidal pulse: (a) in the time domain; (b) in the frequency domain.

### 2.3. Configuration for the Simulations and Experiments

This section describes the configuration of the simulations and experiments. The configuration shown in Figure 7 was applied to the conceptual verification in an anechoic chamber described in Section 2.4, the analysis of operating conditions described in Section 3, and the experimental verification of the proposed method in a classroom described in Section 4.

In Figure 7, A, B, C, and D denote the positions of a silent intruder. Two types of intruders were used in the experiments in an anechoic chamber and a classroom. The first was a PVC pipe 0.3 m in diameter. The second was a person.

The reasons for using two types of intruders were as follows. The PVC pipe was used to identify trends in the localization performance of the proposed active localization method. In other words, using the circular PVC pipe, the reflection sound was uniformly generated even when the sound source was incident at any angle. Therefore, the PVC pipe was used to minimize the change in the absorption ratio of the intruder. The analysis using a PVC pipe was compared with the experimental results of human intrusion and was the background used to simulate the person as a circular boundary.
which corresponded to a center frequency of 1 kHz. This was because when designing the beamformer was a Gaussian-modulated sinusoidal pulse with a 1 kHz center frequency (Equation (10)) and the

According to these results, when reviewing the energy maps again in terms of the maximum error, the proposed concept exhibited no problem under ideal conditions.

In Case 1 (Figure 8a–d), when examining the position estimation of the intruder (i.e., a PVC pipe), although the angle had no error, the error of the distance was observed to reach up to 0.04 m (for position A).

In Case 2 (Figure 8e–h), the error for the angle was confirmed to reach 5° (for position C) and the error for the distance ranged up to 0.13 m (position D) if a person was in each intrusion position. According to these results, when reviewing the energy maps again in terms of the maximum error, Case 1 indicated that the intruder position was estimated with a relatively small error. This was because the PVC pipe had a specific boundary condition at a fixed location without moving. As a result, a consistent reflection wave was measured by the active localization system. However, Case 2

Each superscript on the characters A, B, C, and D of the intruder shows the distance between the active localization system and the intruder position, and each subscript shows the counterclockwise angle between the microphone array and the intruder. The active localization system consisted of a loudspeaker, microphone array, and controller. The positions of the silent intruder were represented by the distance and angle, and the positions of the silent intruder were determined to be the event scenarios close to the wall (positions A and D) or the center of the active localization system (positions B and C).

The size of the control area in the security zone was 2 m × 3 m. The microphones used in the experiment were seven-array microphones. The excitation signal in the simulations and experiments was a Gaussian-modulated sinusoidal pulse with a 1 kHz center frequency (Equation (10)) and the spacing between the microphones was configured to be the same as the Nyquist spacing (λ/2), which corresponded to a center frequency of 1 kHz. This was because when designing the beamformer of the single frequency, the Nyquist spacing had the maximum array gain and directivity [37].

2.4. Preliminary Experiments in Ideal Conditions

This section presents the experimental results in an anechoic chamber. If the proposed method is directly applied in an actual space, exactly matching the analysis with the experimental results becomes difficult because of the various spatial effects (R\text{\scriptsize{\text{\text{reflect}}}}). Therefore, the experimental procedure was performed in an anechoic space to quantitatively verify the accuracy of the proposed approach. In other words, we excluded the environmental elements of the control space and confirmed that the proposed concept exhibited no problem under ideal conditions.

Figure 8 depicts the proposed SRP results obtained from the experiment when a PVC pipe or a person is a silent intruder. Each image shows the intruder position using relative power values (dB).

In Case 1 (Figure 8a–d), when examining the position estimation of the intruder (i.e., a PVC pipe), although the angle had no error, the error of the distance was observed to reach up to 0.04 m (for position A).

In Case 2 (Figure 8e–h), the error for the angle was confirmed to reach 5° (for position C) and the error for the distance ranged up to 0.13 m (position D) if a person was in each intrusion position. According to these results, when reviewing the energy maps again in terms of the maximum error, Case 1 indicated that the intruder position was estimated with a relatively small error. This was because the PVC pipe had a specific boundary condition at a fixed location without moving. As a result, a consistent reflection wave was measured by the active localization system. However, Case 2

![Figure 7](image_url)
indicated that the reflected signals measured by the microphone array were not constant when a person was in the intruder position. The reason was that a slight movement occurred although the person remained in the same position. From this difference, the position estimations of the intruder in the two cases had different results in terms of the maximum error. Nonetheless, we confirmed the feasibility of position estimation through reflections.

Two important conclusions can be drawn. Firstly, the position of a person can be detected using the proposed active localization. Secondly, the energy maps of a person are similar to those of a PVC pipe, which is a circular object. The result indicates that the active localization method can detect the position of an object or a person, and it was the basis for modeling a person as a circular object in the subsequent simulation.

Figure 8. Energy maps of Case 1 and Case 2 for verification of the localization performance in an anechoic chamber: The Case 1 of a PVC pipe in (a) position A; (b) position B; (c) position C; (d) position D; Case 2 of a person in (e) position A; (f) position B; (g) position C; (h) position D.
3. Sound Field Simulation and Its Analysis Using Acoustic Parameters

3.1. Simulation Test for the Reverberant Environment

The active localization method uses reflected sounds; thus, the proposed method is affected by the boundary condition (the property of the wall surface) of the control space. Consequently, the error in Equation (3) increases as the reflection on the wall increases, and the detection performance may be degraded depending on the characteristics of the boundary.

We simulated the environmental operating conditions of the proposed method using the following steps.

STEP 1: The error of localization performance was analyzed by changing the absorption coefficient at the boundary of the target control space (2 m × 3 m).
STEP 2: To examine the correlation between the absorption coefficient of the boundary and the spatial effects, we analyzed the acoustic parameters of the reverberation time (RT20) and early decay time (EDT).
STEP 3: The operating conditions of the active localization were presented using RT20 and EDT.

The experimental approach makes determining sufficient conditions for the proposed method difficult. The results of step 1 based on the finite-difference-time-domain (FDTD) simulation are presented in Section 3.1.2, and the results of steps 2 and 3 are described in Section 3.2.

3.1.1. Simulation Setup

The FDTD method is the numerical solution of the differential equation of a wave. The FDTD method is commonly used for nonstaggered compact schemes expressing only pressure [38] and Yee’s staggered schemes expressing particle velocity and pressure [39].

In this study, the simulation was modeled as Yee’s scheme to use a circular rigid body [40] and a perfectly matched layer (PML) boundary [41]. The circular rigid body boundary was used to model the silent intruder because the characteristics of a person and a PVC pipe were observed to be similar. The PML condition was used to describe the anechoic environment.

The reverberation of the control space was controlled by adjusting the sound absorption coefficient at the boundary. Hence, the momentum equation with the impedance boundary condition was used, and it is expressed as follows:

\[
\begin{align*}
\rho_0 c \frac{1 - \frac{\lambda c}{1 + \lambda c} \zeta}{1 - \frac{\sqrt{1 - \alpha}}{1 - \sqrt{1 - \alpha}}} & = \frac{2\lambda c}{\rho_0 c (1 + \lambda c)} p^{[n]} (u, w) \\
\end{align*}
\]

where \( p \) is the sound pressure; \( v_x \) and \( v_y \) are the particle velocities of the x and y axes, respectively; \( \rho_0 \) is the air density; \( c \) is the speed of sound; \( \lambda \) is the courant number; \( \zeta \) is the specific acoustic impedance; \( \alpha \) is the absorption coefficient; \( n \) is the time index; and \( u \) and \( w \) are indices of the spatial point.

In this study, this impedance boundary condition was derived by combining the asymmetric finite-difference approximation used in [39] and the locally reacting boundary used in a room simulation.
in [38]. The derivation is described in Appendix A. Therefore, we enabled the simulation of the reverberation environment in the Yee scheme using the change in \( \alpha \).

The FDTD simulation utilized a 2 m \( \times \) 3 m control space (Figure 7) and a spatial resolution of 0.01 m. The sampling frequency \( f_{s,FDTD} \) was 49 kHz. As the selection criteria of the parameters, a sampling rate that satisfied the courant condition was selected while the spatial resolution was fixed. The position of the silent intruder was set at representative positions (A, B, C, and D) as mentioned in Section 2.3.

The source model in the FDTD simulation is a physically constrained source (PCS) [42], and the formula is as follows:

\[
p^{[n+1]}(u, w) = p^{[n]}(u, w) + \frac{P_0 e^2 \Delta s}{f_{s,FDTD} \delta s} q^{[n]}(u, w) \tag{16}
\]

\[
q^{[n]}(u, w) = s_p^{[n]} \cdot h_m^{[n]} \tag{17}
\]

\[
s_p^{[n]} = \begin{cases} \omega_c \frac{(2N_p-1)! \sin(n\omega_0)}{b \pi (2N_p+n-1)!!(2N_p-n-1)!!} & \text{if } n = 0 \\ 0 & \text{otherwise} \end{cases} \tag{18}
\]

\[
H_m(e^{i\omega n}) = \frac{b_0 + b_2 e^{-i\omega n}}{1 + a_1 e^{-i\omega n} + a_2 e^{-2i\omega n}} \tag{19}
\]

where \( p^{[n]}(u, w) \) is the pressure node of the source, \( \delta s \) is the spatial resolution, \( A_s = 4\pi a_0^2 \) is the surface area of the sphere in volume velocity, \( q^{[n]}(u, w) \) is the velocity source, \( s_p^{[n]} \) is the maximally flat finite impulse response (FIR) filter, \( h_m^{[n]} \) is the mechanical filter of the source represented by a second-order infinite impulse response (IIR) filter, \( H_m(e^{i\omega n}) \) is \( h_m^{[n]} \) in the frequency domain, \( b_0 = \beta / (M_m \beta^2 + R_m \beta + K_m) \) and \( b_2 = -b_0 \) are the feedforward filter coefficients, \( a_1 = 2(K_m - M_m \beta^2) / (M_m \beta^2 + R_m \beta + K_m) \) and \( a_2 = 1 - 2R_m \beta / (M_m \beta^2 + R_m \beta + K_m) \) are the feedback filter coefficients, \( \beta = \omega_0 / \tan(\omega_0 / 2) \) is the bilinear operator, and \( (\ast) \) denotes the convolution. \( M_m, R_m = M_m \cdot \omega_0 / Q, K_m = M_m \cdot \omega_0^2, \) and \( Q \) are the mass, damping, elasticity, and quality factor constants characterizing the mechanical system of the source, respectively. \( \omega_0 \) is the normalized low resonance frequency of the mechanical system, \( M_p = 4N_p - 1 \) is the FIR filter order, and \( \omega_c \) is the normalized cutoff frequency of the FIR filter.

In this study, \( M_p = 16 \) samples, the normalized cutoff frequency was 0.05, the low resonance frequency was 300 Hz, \( M_m \) was 0.025 Kg, and \( Q \) was set to 0.6.

3.1.2. Simulation Results and Analysis

Figure 9 shows the result images of the active localization method by changing the absorption coefficient of the boundary at position B. The images on the left in Figure 9 show the captured images in the FDTD simulation obtained by reproducing the PCS model. The images on the right indicate the energy maps expressed by the convolution signal of Equation (10) and the impulse response obtained by the FDTD simulation, respectively.

In Figure 9, the reflections propagating from the intruder to the microphone array according to each alpha are similar. The image results show that the magnitude of the wavefront formed by the edge boundary increases as the absorption coefficient of the edge boundary decreases. As a result, the overlap of the reflection formed behind the intruder also increases. In other words, as the reflected sound formed at the boundary becomes significantly louder than the reflected sound produced by the intruder, the spatial effect increases such that the overlapped signal is larger than the intruder’s signal. Therefore, the simulation indicated that the error of position estimation increases with the boundary characteristics of the control space. The simulation results are summarized in Table 1, in which the errors in parentheses represent the angular and distance errors.
Figure 9. Simulation results of the active localization method according to the change in absorption coefficient $\alpha$: (a) $\alpha = 0.9$; (b) $\alpha = 0.7$; (c) $\alpha = 0.5$; (d) $\alpha = 0.3$. The square marker is the actual position and the cross marker is the estimated position.

In Figure 9, the reflections propagating from the intruder to the microphone array according to each alpha are similar. The image results show that the magnitude of the wavefront formed by the edge boundary increases as the absorption coefficient of the edge boundary decreases. As a result, the overlap of the reflection formed behind the intruder also increases. In other words, as the reflected sound formed at the boundary becomes significantly louder than the reflected sound produced by the intruder, the spatial effect increases such that the overlapped signal is larger than the intruder’s signal. Therefore, the simulation indicated that the error of position estimation increases with the
Table 1. Localization performance of the active localization method according to sound absorption at the boundary. The errors in parentheses represent the angular and distance errors.

| A | B | C | D |
|---|---|---|---|
| 135° | 1 m | 90° | 1.5 m | 90° | 2 m | 75° | 2.5 m |
| PML | 135 | (Δθ = 0°) | 1.06 | (Δθ = 0°) | 90 | 1.56 | (Δθ = 0°) | 90 | 2.06 | (Δθ = 0°) | 75 | 2.55 |
| α = 0.9 | 135 | (Δθ = 0°) | 1.07 | (Δθ = 0°) | 90 | 1.48 | (Δθ = 0°) | 90 | 1.98 | (Δθ = 0°) | 75 | 2.47 |
| α = 0.8 | 135 | (Δθ = 0°) | 1.16 | (Δθ = 0°) | 90 | 1.48 | (Δθ = 0°) | 90 | 1.98 | (Δθ = 0°) | 75 | 2.47 |
| α = 0.7 | 135 | (Δθ = 0°) | 1.16 | (Δθ = 0°) | 90 | 1.48 | (Δθ = 0°) | 90 | 2.38 | (Δθ = 0°) | 75 | 2.38 |
| α = 0.6 | 135 | (Δθ = 0°) | 1.24 | (Δθ = 0°) | 90 | 1.48 | (Δθ = 0°) | 90 | 2.38 | (Δθ = 0°) | 75 | 2.38 |
| α = 0.5 | 135 | (Δθ = 0°) | 1.24 | (Δθ = 0°) | 90 | 2.12 | (Δθ = 0°) | 90 | 2.38 | (Δθ = 0°) | 75 | 2.84 |
| α = 0.4 | 135 | (Δθ = 0°) | 1.24 | (Δθ = 0°) | 90 | 2.04 | (Δθ = 0°) | 90 | 2.39 | (Δθ = 0°) | 70 | 2.91 |
| α = 0.3 | 135 | (Δθ = 0°) | 1.24 | (Δθ = 0°) | 90 | 2.04 | (Δθ = 0°) | 90 | 2.30 | (Δθ = 0°) | 70 | 2.91 |

As Table 1 shows, the distance error was affected more by the reflectance of the boundary than by the angular error. There was a 5° error only at the angle at which the sound absorption was below 40% (α × 100%) at the D position. From a distance error point of view, some scenarios failed to detect an intruder. In other words, when the diameter (0.3 m) of the circle considered as the intruder and the predicted distance were combined, the estimated distance exceeded the control space of 2 m × 3 m. The results of α being less than 0.6 at position A and less than 0.5 at position D were the result of detection failure. In addition, when the distance error was viewed in terms of error magnitude, a large error of 0.5 m or more, at α < 0.5 at position B was observed.

Therefore, we confirmed through the simulation that the approach proposed in this paper operates at α ≥ 0.7, for which no angular error exists and the distance error is less than 19%.

In the next section, we describe the relational equation that predicts the environment in which the active localization method operates through the RT20 and EDT of the acoustic parameters. This is because verifying the operation of the proposed method based on the boundary reflectance in a general reverberant environment is very difficult.

3.2. Relationship Analysis of Acoustic Parameters and Absorption Coefficients to Propose Operating Conditions

In this section, the conditions under which the active localization method operates in a reverberant space are explained using the relationship between the acoustic parameters and the absorption coefficient discussed in the previous section.

The proposed approach predicts the position of a silent intruder based on the sound reflected from the intruder, and this phenomenon occurs within a short time; therefore, the pattern of early reflection is very important. If the maximum distance of the active localization system is estimated to be 3 m, the sound source generated by a loudspeaker moves for approximately 17.54 ms when the round-trip distance of the sound source is 6 m and the speed of sound is 342 m/s. In other words, the phenomenon occurring within 18 ms should be analyzed.

Therefore, the EDT and RT20 of the acoustic parameters were used to analyze the control space. EDT includes the direct sound and early reflections, and RT20 has the smallest energy decay time.
considered for the reverberation time indices. EDT and RT20 are expressed in the same equation as follows [43]:

\[
L(t) = 10 \log \frac{\int_0^\infty p^2 dt}{\int_0^\infty p^2 dt}
\]

Equation (20) normalizes the signal power, and we can calculate the time when power decreases from 0 to −10 dB and from −5 to −25 dB through the time variable in the denominator. The time difference of the former is defined as EDT, and the latter is defined as RT20.

When considering the two indices as the early reflection perspective in the RIR, the EDT can physically determine if a large amount of early reflection occurs at the measured location after the direct sound is played. This is because the EDT is the time from the measurement of the direct sound until the signal with an early reflection decrease of −10 dB. RT20 strictly refers to the time when the reverberation energy decreases gradually except for the direct sound and strong early reflection.

To analyze the relationship between the absorption coefficient and EDT/RT20, the microphone was placed at the representative intrusion position shown in Figure 7, and the microphone array signals were compared with the signals of the microphones distributed in the space.

Figure 10a shows the arrangement of the microphones to confirm the operation of the active localization method proposed here. The first to seventh microphones were the array of microphones used in the proposed system, and the eighth to eleventh microphones were placed in the representative positions A, B, C, and D, respectively.

**Figure 10.** Configuration of (a) acoustic parameter tests to verify the active localization method according to the change in absorption coefficient \( \alpha \). (b) Energy decay curve of the ninth microphone.

Figure 10b shows the energy decay curve for the impulse response of the ninth microphone. The energy decay curve of the ninth microphone did not decrease linearly but in a staircase form. This was because the space represented in this simulation was not diffuse. In other words, no diffuse-field reverberation occurred owing to the small space of the simulation and the proximity of the loudspeaker and microphone. As a result, the energy decay curve had an approximate exponential shape of a decay curve, but not the diffuse decay curve (Figure 10b). However, the equation was considered to be suitable for analyzing the space from a physical perspective to confirm the operating conditions of the proposed method. This was because the proposed method was analyzed based on a short time, and the changes in the early reflections were presented by the variation of the EDT and RT20 parameters.

Figure 11 shows the results of EDT and RT20 for each microphone as the sound absorption coefficient decreased. The main point is whether the numerical values measured at the boundaries of.
the control spaces from microphones 1 to 7 and those measured in the control space from microphones 8 to 11 exhibited a specific trend.

![Graph showing early decay time (EDT) and reverberation time (RT20)](image)

**Figure 11.** (a) Early decay time and (b) reverberation time according to change in $\alpha$ based on the finite-difference time-domain (FDTD) simulation.

Figure 11b indicates that the result of the fourth microphone, which was located at the same position as the loudspeaker, was very small compared with the results of other microphones. This was because the loudspeaker and microphone arrangements were very similar such that the characteristics of the room were not sufficiently reflected. Therefore, when analyzing the results of RT20, a criterion for the minimum value to be used for the analysis was necessary.

This criterion was selected as the maximum time for the sound from the loudspeaker to reach the person and back to the microphone again. This is because we can determine that the direct sound and strong early reflection are dominant in a microphone signal if the measured time of RT20 is shorter than the propagation time of the sound source generated by the loudspeaker.

The farthest distance in the configuration of this study was 2.62 m, which was the distance from microphone 4 to the upper corner (2.92 m) minus the distance of 0.3 m at which a person can stand. The criterion time can be selected as follows:

$$t_c = \frac{2d_{\text{max}}}{c} \times 100 = \frac{2 \times 2.62}{342} \times 100 = 15.3 \text{ ms}$$

where $t_c$ is the criterion time, $d_{\text{max}}$ is the maximum distance of a sound source in the control domain, and $c$ is the speed of sound.

Therefore, when analyzing RT20, values less than $t_c$ were excluded from the analysis.

Figure 12 is a graph showing the minimum, maximum, and median values of EDT and RT20 in the microphone array and control space according to $\alpha$. In this scenario, the microphone signal that did not satisfy $t_c$ was excluded from the RT20 analysis. The results of the microphones in the array and control space are represented by the red dashed and blue solid lines, respectively. The marker on each graph is the median value, the top line of the deviation is the maximum value, and the bottom line is the minimum value.

The EDT results shown in Figure 12a indicate that the median value of the microphones in the control space was higher than that in the array. However, the deviation confirmed that the EDT results in the array were large depending on the absorption coefficient.

The RT20 results depicted in Figure 12b indicate that until $\alpha = 0.7$, the median value of the control space was larger than that of the array, but from 0.6, the opposite result was observed. The deviation tended to increase and decrease as $\alpha$ decreased.
Figure 11. (a) Early decay time and (b) reverberation time according to change in $\alpha$ based on the finite-difference time-domain (FDTD) simulation.

Figure 11b indicates that the result of the fourth microphone, which was located at the same position as the loudspeaker, was very small compared with the results of other microphones. This was because the loudspeaker and microphone arrangements were very similar such that the characteristics of the room were not sufficiently reflected. Therefore, when analyzing the results of RT20, a criterion for the minimum value to be used for the analysis was necessary.

This criterion was selected as the maximum time for the sound from the loudspeaker to reach the person and back to the microphone again. This is because we can determine that the direct sound and strong early reflection are dominant in a microphone signal if the measured time of RT20 is shorter than the propagation time of the sound source generated by the loudspeaker.

The farthest distance in the configuration of this study was 2.62 m, which was the distance from microphone 4 to the upper corner (2.92 m) minus the distance of 0.3 m at which a person can stand. The criterion time can be selected as follows:

$$t_{c} = \max_{m} \left( \sqrt{d_{m}^2} \right)$$

where $d_{m}$ is the maximum distance of a sound source in the control domain, and $c$ is the speed of sound.

Therefore, when analyzing RT20, values less than $t_{c}$ were excluded from the analysis.

Figure 12 is a graph showing the minimum, maximum, and median values of EDT and RT20 in the microphone array and control space according to $\alpha$. In this scenario, the microphone signal that did not satisfy $t_{c}$ was excluded from the RT20 analysis. The results of the microphones in the array and control space are represented by the red dashed and blue solid lines, respectively. The marker on each graph is the median value, the top line of the deviation is the maximum value, and the bottom line is the minimum value.

Analyzing the values in Figure 12 according to the conclusion in Section 3.1.2 that the proposed approach operated in an environment with $\alpha > 0.7$, the following features were obtained. From the EDT results in Figure 12a, we observed that the maximum values of the array became smaller than the maximum values of the control space when $\alpha$ was greater than 0.7. When the results of RT20 in Figure 12b were analyzed as a median value, when alpha is greater than 0.7, the median values of the array were smaller than those of the control space. The results are summarized in Table 2.

Table 2. Simulation results of acoustic parameters to confirm the operating conditions of the active localization method.

| $\alpha$ | EDT 1 (ms) Max Value | RT20 2 (ms) Median Value |
|----------|----------------------|--------------------------|
|          | in a Linear Array    | in Control Space         | in a Linear Array    | in Control Space |
| 0.9      | 0.53                 | 3.12                     | 15.57                | 17.36            |
| 0.8      | 3.10                 | 5.42                     | 16.73                | 17.66            |
| 0.7      | 6.04                 | 10.75                    | 17.89                | 21.23            |
| 0.6      | 14.83                | 12.16                    | 33.34                | 29.54            |
| 0.5      | 16.53                | 13.36                    | 34.02                | 30.33            |
| 0.4      | 17.81                | 18.44                    | 50.32                | 43.13            |
| 0.3      | 21.44                | 24.97                    | 60.53                | 56.14            |

1 EDT: Early decay time (EDT) (0 to −10 dB), 2 RT20: Reverberation time (RT20) (−5 to −25 dB).

As the results in Tables 1 and 2 show, the active localization method proposed in this paper can detect the position of a person and an object under the following conditions:

$$\max_{m} \left[ \text{EDT}_{m}^{\text{array}} \right] < \max_{m} \left[ \text{EDT}_{m}^{\text{spatial}} \right]$$

$$\text{median} \left[ \text{RT20}_{m}^{\text{array}} \right] < \text{median} \left[ \text{RT20}_{m}^{\text{spatial}} \right]$$

$$\begin{cases} \text{RT20}_{m}^{\text{array}} > t_c \\ \text{RT20}_{m}^{\text{spatial}} > t_c \end{cases}$$

where $m$ is the microphone index and $t_c$ is the criterion time.

Equation (22) indicates a condition in which the maximum EDT value of the array is smaller than that of the control space. Equation (23) indicates that the median value of RT20 in the array is less than its median value in the control space, where RT20 values above $t_c$ are used.

Therefore, we observed that if the microphones are installed in the array and control space, the acoustic parameters of EDT and RT20 satisfy the conditions of Equations (22) and (23), and the active localization method can be implemented.
4. Experimental Results of Active Localization in a Reverberant Environment

This section presents the experimental results to verify the proposed method. In Section 2.4, we confirmed the feasibility of the proposed approach in an anechoic chamber, that is, the concept of detecting the position of a person or an object through a reflected sound. The results of an anechoic chamber indicated that there was no error in Equation (3). However, in an actual space in which reverberation exists, an error occurs in Equation (3). Therefore, the conditions under which the active localization method can operate in the reverberation space are identified in Section 3.

We used Equations (22) and (23) to predict whether the active localization method would function in a classroom, and we describe the experimental results using the proposed method to estimate the position of a PVC pipe and a person.

4.1. Experimental Configuration and Operating Conditions Test

Figure 13 shows the experimental environment of an empty classroom. The experiments were performed at the same position as the silent intruder (Figure 7). The room acoustic parameters were measured using the configuration shown in Figure 10a, and the results are presented in Table 3.

Table 3 shows the EDT and RT20 measured at seven microphones in an array, and the EDT and RT20 measured at the eighth to eleventh microphones as the representative intrusion shown in Figures 7 and 10a.
Table 3. Results of room acoustic parameters measured in the control space as in Figure 13 at positions shown in Figure 10a.

| Position          | EDT (ms) | RT20 (ms) |
|-------------------|----------|-----------|
| 1                 | 8.1      | 24.0      |
| 2                 | 2.7      | 20.0      |
| 3                 | 2.0      | 11.6      |
| 4                 | 2.0      | 5         |
| 5                 | 2.1      | 15.5      |
| 6                 | 2.6      | 20.1      |
| 7                 | 9.0      | 23.5      |
| 8 (A)             | 10.5     | 22.7      |
| In control space  |          |           |
| 9 (B)             | 22.8     | 23.5      |
| 10 (C)            | 13.2     | 23.6      |
| 11 (D)            | 12.2     | 25.3      |

Firstly, when ascertaining the operating conditions using the EDT of Equation (22), the maximum value measured in the microphone array was 9.0 ms and the maximum value measured in the control space was 22.8 ms. Therefore, we confirmed that Equation (22) was satisfied.

Secondly, when the operating condition using the median value of RT20 in Equation (23) was applied to the data in Table 3, the median value of the array was 20.1 ms. By excluding the RT20 that did not satisfy Equation (24), the median value of the distributed microphones in the control space was 23.5 ms. Therefore, we confirmed that Equation (23) was also satisfied.

The results indicate that the proposed active localization method operates even if reverberation exists in the control space set as the security space. The localization results using SRP energy maps are discussed in the following section.

4.2. Localization Performance in a Reverberant Environment

Figure 14 depicts the energy maps obtained from the experimental results. Case 1 shows the test results when the PVC pipe was considered as a silent intruder, and Case 2 shows the results when a person is the silent intruder. Each image shows the intruder position using relative power values (dB). The square marker is the actual position and the cross marker is the estimated position.

To analyze the experimental results in Figure 14, we compared the estimated position results with those in Table 1, which lists the simulation results of the reverberation environment. In Table 1, when examining the results of $\alpha$ greater than 0.7, which is the range in which the active localization method operates, no error of angle was observed and the error of distance was up to 19% (distance error 0.38 m).

The experimental results of the reverberation environment in Figure 14 indicate that the angle had no error, and the error for distance was within 6.5% (distance error 0.13 m). Therefore, the proposed active localization method can be implemented if the operating conditions of Equations (22) and (23) are satisfied, as discussed in Section 3.2. However, the position detection results of Case 2 shown in Figure 14 indicate an increased error compared with the results of the PVC pipe. To analyze this, the results in both an anechoic chamber and a classroom are summarized quantitatively in Table 4 as the error between the actual and estimated values of each experimental configuration. These position errors represent angle and distance errors. The results of the localization performances are compared in terms of the type of silent intruder (a PVC pipe or a person).
Table 3 shows the EDT and RT20 measured at seven microphones in an array, and the EDT and RT20 measured at the eighth to eleventh microphones as the representative intrusion shown in Figures 7 and 10a.

Firstly, when ascertaining the operating conditions using the EDT of Equation (22), the maximum value measured in the microphone array was 9.0 ms and the maximum value measured in the control space was 22.8 ms. Therefore, we confirmed that Equation (22) was satisfied.

Secondly, when the operating condition using the median value of RT20 in Equation (23) was applied to the data in Table 3, the median value of the array was 20.1 ms. By excluding the RT20 that did not satisfy Equation (24), the median value of the distributed microphones in the control space was 23.5 ms. Therefore, we confirmed that Equation (23) was also satisfied.

The results indicate that the proposed active localization method operates even if reverberation exists in the control space set as the security space. The localization results using SRP energy maps are discussed in the following section.

4.2. Localization Performance in a Reverberant Environment

Figure 14 depicts the energy maps obtained from the experimental results. Case 1 shows the test results when the PVC pipe was considered as a silent intruder, and Case 2 shows the results when a person is the silent intruder. Each image shows the intruder position using relative power values (dB). The square marker is the actual position and the cross marker is the estimated position.

Figure 14. Energy maps of Case 1 and Case 2 to verify the localization performance in a classroom: The Case 1 of a PVC pipe in (a) position A; (b) position B; (c) position C; (d) position D; Case 2 of a person in (e) position A; (f) position B; (g) position C; (h) position D. The square marker is the actual position and the cross marker is the estimated position.

Table 4. Position errors of a PVC pipe and a person in terms of localization performance.

| Position | PVC Pipe | Person |
|----------|----------|--------|
|          | A1_{135} | B^{1.5}_{90} | C^{2}_{90} | D^{2.5}_{75} |
| A1_{135} | \( \Delta \theta = 0^\circ \) \( \Delta r = 0.04 \text{ m} \) & \( \Delta \theta = 0^\circ \) \( \Delta r = 0.06 \text{ m} \) & \( \Delta \theta = 0^\circ \) \( \Delta r = 0.03 \text{ m} \) & \( \Delta \theta = 0^\circ \) \( \Delta r = 0.03 \text{ m} \) |
|          | (4%)     | (6%)   | (3%)  | (1.6%)   |
| B^{1.5}_{90} | \( \Delta \theta = 0^\circ \) \( \Delta r = 0.02 \text{ m} \) & \( \Delta \theta = 0^\circ \) \( \Delta r = 0.03 \text{ m} \) & \( \Delta \theta = 0^\circ \) \( \Delta r = 0.09 \text{ m} \) & \( \Delta \theta = 0^\circ \) \( \Delta r = 0.11 \text{ m} \) |
|          | (1.33%)  | (2%)   | (6%)  | (34%)    |
| C^{2}_{90} | \( \Delta \theta = 0^\circ \) \( \Delta r = 0.03 \text{ m} \) & \( \Delta \theta = 0^\circ \) \( \Delta r = 0.13 \text{ m} \) & \( \Delta \theta = 0^\circ \) \( \Delta r = 0.05 \text{ m} \) & \( \Delta \theta = 0^\circ \) \( \Delta r = 0.43 \text{ m} \) |
|          | (1.5%)   | (6.5%) | (2.5%)| (21.5%)  |
| D^{2.5}_{75} | \( \Delta \theta = 0^\circ \) \( \Delta r = 0.01 \text{ m} \) & \( \Delta \theta = 0^\circ \) \( \Delta r = 0.04 \text{ m} \) & \( \Delta \theta = 0^\circ \) \( \Delta r = 0.13 \text{ m} \) & \( \Delta \theta = 0^\circ \) \( \Delta r = 0.27 \text{ m} \) |
|          | (0.4%)   | (1.6%) | (5.2%)| (10.8%)  |
The data of the anechoic chamber indicated the initial error of the proposed method under the condition that no effect of reflection and reverberation occurred in the control space, and the data of the classroom indicated the performance of the proposed method under conditions of reflection and reverberation. In an anechoic environment that represented the initial error, the position error increased in the scenario of a person compared with that of a PVC pipe. This was caused by the slight movement of the person, and the data results in Table 4 indicate that the position error can be further increased when this movement is combined with a reverberation environment.

From the experimental results of Case 2 in a classroom, we confirmed that the results of the PVC pipe had a small error, within 6.5%, owing to the nonmovement of a pipe, whereas the cases of a human intruder indicated a relatively large error, within 5° of the estimated angle and 34% of the estimated distance.

Therefore, the above results of localization performance improve the limitations of the existing acoustic-based security system, for which an intruder must generate sound. Moreover, the proposed method estimates the x and y positions using a linear microphone array in a two-dimensional security space.

5. Conclusions and Discussion

In this paper, a new active localization method is proposed to estimate the position of a silent intruder.

For feasibility testing and analysis of the proposed method, we performed the following four steps. Firstly, feasibility tests were performed in an anechoic chamber. Secondly, an FDTD simulation was conducted to verify that the proposed method operates according to the reflection in the boundary of the control space. Thirdly, EDT and RT20 were used to represent the conditions under which active localization can operate in a reverberant environment through FDTD simulation data. Finally, the operation of the active localization method in a classroom was confirmed under conditions based on the EDT and RT20, and then we analyzed the localization results of a PVC pipe and a person through energy maps. Therefore, the proposed method was verified for the position estimation of a silent intruder. The active localization method is expected to be applied in home security systems in conjunction with conventional security sensors to improve the capability of intrusion detection because the proposed system can estimate the position of a silent intruder and can be implemented using loudspeakers and microphones built-in in home appliances.

In a further study, we intend to expand the frequency band to conduct more precise analyses of the security space, represent the SRP energy maps using wideband data, and design digital filters to determine the robustness of the proposed method.

Author Contributions: Conceptualization, K.K., S.W., and S.Q.L.; methodology, K.K.; software, K.K.; validation, K.K., H.R., S.W., and S.Q.L.; formal analysis, K.K.; data curation, K.K., and H.R.; writing—original draft preparation, K.K.; writing—review and editing, S.W.; supervision, S.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the “GIST Research Institute (GRI)” grant funded by the GIST in 2020.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The simulation was modeled as the Yee scheme of the FDTD method (Figure A1).

The wave equation is expressed as a two-dimensional linear acoustic domain [39].

\[
v^\text{[n+0.5]}_x(u, w + 0.5) = v^\text{[n-0.5]}_x(u, 0.5, w) - \frac{\Delta t}{\rho_0 \delta s} \times [p^\text{[n]}(u + 1, w) - p^\text{[n]}(u, w)] \tag{A1}
\]

\[
v^\text{[n+0.5]}_y(u, w + 0.5) = v^\text{[n-0.5]}_y(u, 0.5, w) - \frac{\Delta t}{\rho_0 \delta s} \times [p^\text{[n]}(u, w + 1) - p^\text{[n]}(u, w)] \tag{A2}
\]
\[ p^{n+1}(u, w) = p^n(u, w) - \frac{\rho_0 c^2 \delta t}{\delta s} \left[ v_x^{n+0.5}(u + 0.5, w) - v_x^{n+0.5}(u - 0.5, w) \right] - \frac{\rho_0 c^2 \delta t}{\delta s} \left[ v_y^{n+0.5}(u, w + 0.5) - v_y^{n+0.5}(u, w - 0.5) \right] \]  

where \( p \) is the sound pressure, \( v_x \) and \( v_y \) are the particle velocities of the x and y axes, respectively, \( \delta s \) is the spatial discretization step, \( \delta t \) is the time discretization step, \( u \) and \( w \) are indices of the spatial point, \( \rho_0 \) is the air density, and \( c \) is the speed of sound.

The equation of the boundary condition causing reflection is expressed using the asymmetric finite-difference approximation in [39].

\[ \frac{\partial p^n}{\partial n}(u + 0.5, w) = \frac{2}{\delta s} \left[ p^n(u + 0.5, w) - p^n(u, w) \right] \]  

In Figure A1, the boundary of the right side on the x-axis is expressed in Equation (A5). No term of \( p^n(u + 1, w) \) exists in the momentum Equation (A1). Therefore, when the approximation of Equation (A4) is introduced into Equation (A1), it is expressed as follows:

\[ v_x^{n+0.5}(u + 0.5, w) = v_x^{n-0.5}(u + 0.5, w) - \frac{2\delta t}{\rho_0 \delta s} \times [p^n(u + 0.5, w) - p^n(u, w)] \]  

In Equation (A5), \( p^n(u + 0.5, w) \) represents the velocity point in the x direction as a spatial point that does not exist in Figure A1. The point \( (u + 0.5, w) \) is an impedance boundary of the FDTD domain, which is suitable for expressing the locally reacting boundary affected only by a normal velocity because of the lattice structure of the Yee scheme. Equation (A6) represents the acoustic impedance with the locally reacting boundary.

\[ Z = \frac{P}{v_n} \]  

where \( P \) is the acoustic pressure, and \( v_n \) is the normal velocity.

\[ p^n(u + 0.5, w) \] in (29) is replaced with \( Zv^n(u + 0.5, w) \) based on Equation (A6). The linear interpolation to represent \( v_x^n \) as \( v_x^{n-0.5} \) and \( v_x^{n+0.5} \) is as follows:

\[ v_x^{n+0.5}(u + 0.5, w) = v_x^{n-0.5}(u + 0.5, w) - \frac{2\delta t}{\rho_0 \delta s} \times [Zv^n(u + 0.5, w) - p^n(u, w)] \]

\[ v_x^{n+0.5}(u + 0.5, w) = v_x^{n-0.5}(u + 0.5, w) - \frac{2\delta t}{\rho_0 \delta s} \left[ \frac{Z}{2} [v_x^{n-0.5}(u + 0.5, w) + v_x^{n+0.5}(u + 0.5, w)] - p^n(u, w) \right] \]  

The following concept is introduced to assign the frequency-independent absorption coefficient to Equation (A7).
In [44], the wall impedance is frequently divided by the characteristic impedance of air. The resulting quantity, expressed as Equation (A8), is called the specific acoustic impedance.

\[ \zeta = \frac{Z}{\rho_0 c} \]  

(A8)

where \( Z \) is the acoustic impedance, \( \rho_0 \) is the density of air, and \( c \) is the speed of sound.

The specific acoustic impedance is also represented by the reflection coefficient \( R \).

\[ \zeta = \frac{1 + R}{1 - R} \]  

(A9)

The intensity of a plane wave is proportional to the square of the pressure amplitude. Therefore, the intensity of the reflected wave is smaller by a factor \( |R|^2 \) than that of the incident wave. This quantity is called the “absorption coefficient” of the wall.

\[ \alpha = 1 - |R|^2 \]  

(A10)

Therefore, the momentum equation of (A7) in the boundary is rewritten as

\[ v_x^{[n+0.5]}(u + 0.5, w) = \left( \frac{1 - \lambda_c \zeta}{1 + \lambda_c \zeta} \right) v_x^{[n-0.5]}(u + 0.5, w) + \frac{2\lambda_c}{\rho_0 c(1 + \lambda_c \zeta)} p^{[n]}(u, w) \]  

(A11)

where \( \lambda_c \) is the Courant number. \( \zeta \) is the specific acoustic impedance and \( \alpha \) is the function of the absorption coefficients of Equations (A9) and (A10).

Appendix B

Figure A2 shows feasibility experiments in which the hidden intruder can be determined with the sound field variation proposed in this paper.

**Figure A2.** Experimental setup and results of a hidden object and a hidden person: (a) Experimental configuration of the detection performance; (b) An experimental picture; Energy maps of (c) a hidden PVC pipe and (d) a hidden person.
From the perspective of position estimation, a large position error of 0.5 m or more is generated for the distance, but from the perspective of detection, it exhibits a meaningful result that a person or object hidden behind an obstacle can be detected.

The echolocation method using the audible frequency proposed in this paper has a functional advantage that even a hidden person can be detected through the amount of scattering variation.

References
1. Ding, D.; Cooper, R.A.; Pasquina, P.F.; Fici-Pasquina, L. Sensor technology for smart homes. *Maturitas* 2011, 69, 131–136. [CrossRef]
2. Ueda, K.; Suwa, H.; Arakawa, Y.; Yasumoto, K. Exploring Accuracy-Cost Tradeoff in In-Home Living Activity Recognition based on Power Consumptions and User Positions. In Proceedings of the IEEE International Conference on Computer and Information Technology, Liverpool, UK, 21–23 September 2015; pp. 1130–1137.
3. Laout, G.; Zhang, Y.; Harrison, C. Synthesis Sensors: Toward General-Purpose Sensing. In Proceedings of the CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 3986–3999.
4. Chilipirea, C.; Ursache, A.; Popa, D.O.; Pop, F. Energy efficiency robustness for IoT: Building a smart home security. In Proceedings of the 2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 8–10 September 2016; pp. 43–48.
5. Dahmen, J.; Cook, D.J.; Wang, X.; Honglei, W. Smart secure homes: A survey of smart home technologies that sense, assess, and respond to security threats. *J. Reliab. Intell. Environ.* 2017, 3, 83–98. [CrossRef]
6. Itai, A.; Yasukawa, H. Personal Identification using Footstep based on Wavelets. In Proceedings of the 2006 International Symposium on Intelligent SignalProcessing and Communications, Tottori, Japan, 12–15 December 2006; pp. 383–386.
7. Olalekan, O.B.; Toluwani, O.V. Automated Home Security System: A Review. *MAYFEB J. Electr. Electron. Eng.* 2016, 1, 7–16.
8. Crocco, M.; Cristani, M.; Trucco, A.; Murino, V. Audio surveillance: A systematic review. *ACM Comput. Surv.* 2016, 48, 52:1–52:46. [CrossRef]
9. Foggia, P.; Petkov, N.; Saggese, A.; Strisciuglio, N.; Vento, M. Reliable detection of audio events in highly noisy environments. *Pattern Recognit. Lett.* 2015, 65, 22–28. [CrossRef]
10. Jung, K.K.; Shin, H.S.; Kang, S.H.; Eom, K.H. Object tracking for security monitoring system using microphone array. In Proceedings of the International Conference on Control, Automation and Systems, Seoul, Korea, 17–20 October 2007; pp. 2351–2354.
11. Dostalek, P.; Vasek, V.; Kresalek, V.; Navratil, M. Utilization of audio source localization in security systems. In Proceedings of the 43rd International Conference on Security Technology, Zurich, Switzerland, 5–8 October 2009; pp. 305–311.
12. Transfield, P.; Martens, U.; Binder, H.; Schypior, T.; Fingscheidt, T. Acoustic event source localization for surveillance in reverberant environments supported by an event onset detection. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015; pp. 2629–2633.
13. Abu-El-Quran, A.R.; Goubran, R.A.; Chan, A.D.C. Security monitoring using microphone arrays and audio classification. *IEEE Trans. Instrum. Meas.* 2006, 55, 1025–1032. [CrossRef]
14. Kawamoto, M.; Asano, F.; Kurumatani, K.; Hua, Y. A system for detecting unusual sounds from sound environment observed by microphone arrays. In Proceedings of the 5th International Conference on Information Assurance and Security, Xi’An, China, 18–20 August 2009; pp. 729–732.
15. Chen, B.W.; Chen, C.Y.; Wang, J.F. Smart Homecare Surveillance System: Behavior Identification Based on State-Transition Support Vector Machines and Sound Directivity Pattern Analysis. *IEEE Trans. Syst. Man Cybern. Syst.* 2013, 43, 1279–1289. [CrossRef]
16. Choi, Y.K.; Kim, K.M.; Jung, J.W.; Chun, S.Y.; Park, K.S. Acoustic intruder detection system for home security. *IEEE Trans. Consum. Electron.* 2005, 51, 130–138. [CrossRef]
17. Lee, S.Q.; Park, K.H.; Kim, K.; Ryu, H.M.; Wang, S. Intrusion detection based on the sound field variation in audible frequency-general sound space case. In Proceedings of the 19th International Conference on Sound and Vibration (ICSV), Vilnius, Lithuania, 8–12 July 2012; pp. 1–8.
18. Lee, C.; Kim, D.; Kim, K. Acoustic detection based on coherence bandwidth. *Electron. Lett.* 2015, 51, 1387–1388. [CrossRef]

19. Ishigaki, T.; Higuchi, T.; Watanabe, K. An Information Fusion-Based Multiobjective Security System with a Multiple-Input/Single-Output Sensor. *IEEE Sens. J.* 2007, 7, 734–742. [CrossRef]

20. Dhake, P.S.; Borde, S.S. Embedded Surveillance System Using PIR Sensor. *Int. J. Adv. Technol. Eng. Sci.* 2014, 2, 31–36.

21. Sonbul, O.; Kalashnikov, A.N. Low Cost Ultrasonic Wireless Distributed Security System for Intrusion Detection. In Proceedings of the 7th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), Berlin, Germany, 12–14 September 2013; pp. 235–238.

22. Lee, Y.; Han, D.K.; Ko, H. Acoustic Signal Based Abnormal Event Detection in Indoor Environment using Multiclass Adaboost. *IEEE Trans. Consum. Electron.* 2013, 59, 615–622. [CrossRef]

23. Kuc, R. Biomimetic Sonar Locates and Recognizes Objects. *IEEE J. Ocean. Eng.* 1997, 22, 616–624. [CrossRef]

24. Reijniers, J.; Peremans, H. Biomimetic Sonar System Performing Spectrum-Based Localization. *IEEE Trans. Robot.* 2007, 22, 1151–1159. [CrossRef]

25. Sunil, B.H. Household Security System Based on Ultrasonic Sensor Technology with SMS Notification. *Eur. J. Acad. Essays* 2014, 1, 6–9.

26. Sharma, R.; Dhingra, S.K.; Pandey, N.; Garg, R.; Singhal, R. Electric Field and Ultrasonic Sensor Based Security System. In Proceedings of the 2010 International Conference on Intelligent System, Modeling and Simulation, Liverpool, UK, 28–30 July 2010; pp. 423–426.

27. Kim, K.; Kim, D.; Ryu, H.; Wang, S.; Lee, S.Q.; Park, K.H. Active localization of a silent intruder with audible frequency in 2D security space. In Proceedings of the International Conference and Exposition on Noise Control Engineering, New York, NY, USA, 19–22 August 2012; pp. 1–8.

28. Dokmanic, I.; Parhizkar, R.; Walther, A.; Lu, M.Y.; Veterli, M. Acoustic echoes reveal room shape. *Proc. Natl. Acad. Sci. USA* 2013, 110, 12186–12191. [CrossRef]

29. Dokmanic, I.; Parhizkar, R.; Ranieri, J.; Vetterli, M. Euclidean Distance Matrices. *IEEE Signal Process. Mag.* 2015, 32, 12–30. [CrossRef]

30. Krekovic, M.; Dokmanic, I.; Vetterli, M. EchoSLAM: Simultaneous Localization and Mapping with Acoustic Echoes. In Proceedings of the 2016 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 11–15.

31. Kim, Y.H. Sound Propagation: An Impedance Based Approach; John Wiley & Sons (Asia) Pte Ltd., Clementi Loop, #02-01: Singapore, 2010; p. 139.

32. DiBiase, J.H.; Silverman, H.F.; Brandstein, M.S. Robust Localization in Reverberant Rooms. In *Microphone Arrays: Signal Processing Techniques and Applications*; Brandstein, M., Ward, D., Eds.; Springer: Berlin, Germany, 2001; pp. 157–180.

33. Zhang, C.; Florencio, D.; Zhang, Z. Why does PHAT work well in low noise, reverberative environments? In Proceedings of the 2008 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), Las Vegas, NV, USA, 31 March–4 April 2008; pp. 2565–2568.

34. Trees, H.L.V. *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2002; p. 66.
40. Tornberg, A.K.; Engquist, B. Consistent boundary conditions for the Yee scheme. *J. Comput. Phys.* 2008, 227, 6922–6943. [CrossRef]

41. Yuan, X.; Broup, D.; Wiskin, J.W.; Berggren, M.; Eidens, R.; Johnson, S.A. Formulation and Validation of Berenger’s PML absorbing Boundary for the FDTD Simulation of Acoustic Scattering. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 1997, 44, 816–822. [CrossRef]

42. Sheaffer, J.; Walstijn, M.V.; Fazenda, B. Physical and numerical constraints in source modeling for finite difference simulation of room acoustics. *J. Acoust. Soc. Am.* 2014, 135, 251–261. [CrossRef]

43. Hak, C.C.J.M.; Wenmaekers, R.H.C.; van Luxemburg, L.C.J. Measuring Room Impulse Responses: Impact of the Decay Range on Derived Room Acoustic Parameters. *Acta Acust. United Acust.* 2012, 98, 907–915. [CrossRef]

44. Kuttruff, H. *Room Acoustics*, 4th ed.; Spon Press: London, UK, 2000; p. 33.

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 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/).