Quality and User Experience (2024) 9:2
https://doi.org/10.1007/s41233-024-00066-x

RESEARCH ARTICLE

Recognizing object localization using acoustic markers with active acoustic sensing

Fuma Kishi[1] · Kodai Ito[2] · Kazuyuki Fujita[3] · Yuichi Itoh[2]

Received: 31 May 2023 / Published online: 11 March 2024

Abstract
We propose a low-cost sensing system that recognizes an object's location on a surface using active acoustic sensing. Our proposed system uses a thin speaker attached to an object as a marker and estimates the marker's location from the sound source. Localization is achieved through machine learning (random forest) based on the property that high-frequency components of sound decrease more than low-frequency components with distance. We additionally implemented a system to simulate the condition where multiple objects are placed simultaneously and to estimate the frequency response of those objects from training data where only a single object is placed. Performance tests show that our system localizes a single object with a mean absolute error of 0.41 cm in a 20 cm square area on a wooden deck and also localizes the placement of four objects with an accuracy of 1.83 cm while saving 83.3% of the effort needed to collect the training data.

Keywords Tangible user interface · Active acoustic sensing · Sound localization · Machine learning

Introduction

There has been much research on computer recognition of information in recent years, such as the type and position of real-world objects and interactions with them. For example, in Project Zanzibar, the computer recognizes the type and position of a remarkable toy figure so that it can be played with using computer images [1]. Prior research efforts considered using various types of sensors for recognition of object information, such as cameras (e.g., [2]) and pressure sensors (e.g., [3]). However, camera-based methods often require placing the camera at a high location such as the ceiling to eliminate occlusion and pressure sensors need to be spread over the sensor surface, resulting in high prototyping costs for object information recognition. Therefore, it is difficult for developers to make existing, everyday objects interactive.

As an approach to address this limitation, we focus on active acoustic sensing. We can achieve this sensing technique by propagating acoustic signals between a microphone and a loudspeaker and analyzing their responses. This low-cost system can be easily implemented by simply attaching a microphone and speaker to an object or surface. In the active acoustic sensing method, a thin speaker is attached to an object as an acoustic monitor and its surface becomes the sensing surface. Each acoustic marker emits an acoustic signal of a specific frequency assigned to its corresponding object. The type and location of an object on the surface can then be estimated by the propagated acoustic signals.

In this case, analyzing the acoustic signal and estimating the location of marker-attached objects is the same as localizing the source of an acoustic signal emitted from the

Kodai Ito, Kazuyuki Fujita and Yuichi Itoh have contributed equally to this work.

✉ Yuichi Itoh
itoh@it.aoyama.ac.jp

Fuma Kishi
fuma.kishi@x-lab.team

Kodai Ito
kodai.ito@it.aoyama.ac.jp

Kazuyuki Fujita
k-fujita@riec.tohoku.ac.jp

[1] Graduate School of Information Science and Technology, Osaka University, 2-1, Yamadaoka, Suita, Osaka 565-0871, Japan

[2] College of Science and Engineering, Aoyama Gakuin University, 5-10-1, Fuchinobe, Chuo-ku, Sagamihara, Kanagawa 252-5258, Japan

[3] Research Institute of Electrical Communication, Tohoku University, 6-6 Aoba, Aramaki, Aoba-ku, Sendai, Miyagi 980-8579, Japan

marker. One promising method for sound source localization is time difference of sound arrivals (TdoA) positioning [4], which uses an array of microphones. This method uses multiple microphones to sense acoustic signals and localizes the sound source based on the arrival time difference of acoustic signals to each microphone. However, its distance resolution depends on the sampling rate of the audio hardware. Therefore, a sampling rate of 1 MHz or higher is often required for localizing a sound source to a few cms. Audio measurement devices thus require high hardware performance and are not suitable for sensing in a small area, such as the size of a desk.

In this study, we propose a novel system of active acoustic sensing based on the characteristic that high-frequency components of sound damp out with distance, therefore enabling sensing in a small area without depending on the performance of audio hardware such as sampling rate. In addition, as a more advanced recognition method, we propose a system for estimating the training data of multiple objects placed simultaneously from the training data of individual objects placed singularly. This estimation is achieved by applying the principle of superposition to sound waves. We expect this system to significantly reduce the time and effort required to collect the training data. In addition, by separating the frequencies of the signals emitted from the acoustic markers, we attempt to create a sound source localization system in which the acoustic signals do not interfere with each other.

This paper first describes the implementation of our prototype device consisting of a microphone, a piezoelectric element as a speaker, and an audio interface. Next, we describe the localization of a single object on a wooden desk in two dimensions and evaluate the estimation error and learning cost. Then, we describe a multi-object recognition system based on the principle of wave superposition and show an estimation test of four different objects simultaneously. Finally, based on these results, we discuss our system and possible applications.

Our main contributions are: (i) We proposed and tested a marker system for recognizing object placement on a surface using active acoustic sensing and (ii) our additional evaluation confirmed the possibility of simultaneously estimating the placement of four objects.

Related work

Object recognition

A major approach to object recognition involves camera-based methods, such as using images with neural networks [2]. However, optical occlusion is a problem for these methods by interrupting sensing each time an object is manipulated. To solve this, other studies have attempted to place a camera or projector under the sensing surface [5] to analyze the bottom shape of the object [6] and estimate the type and location of the object. However, a table with a camera and projector is inevitably a large device. Another approach is based on the physical characteristics of the object. Studies have been reported on recognizing objects by using the electromagnetic noise of electronic devices [7] and the electrical conductivity of objects [8]. However, these methods are limited in that they can only recognize particular objects. In addition, a method to identify objects from their footprints using Fourier-transform infrared spectroscopy (FTIR; pressure-sensitive sheeting) on the floor has been investigated [9]. However, it is difficult to distinguish objects with similar footprints.

In summary, there are various approaches to object recognition. However, these approaches are limited in the objects that can be detected and the implementing cost of these systems is high because a sensor needs to be embedded in the sensing surface. We therefore employ acoustic sensing for object recognition as it easily recognizes existing object information at low cost.

Acoustic sensing

Passive acoustic sensing

Passive acoustic sensing detects and analyzes the sound generated by tapping and scraping sounds without any applying signal. It can recognize touch detection [10, 11], swipe gesture detection [12, 13], objects used for touch [14, 15], and human activities (e.g., using an oven, turning on a faucet) [16–19].

Acoustic Pulse Recognition (APR) is one of the well-known passive acoustic sensing techniques. This method identifies the touched location by matching the pattern of the acoustic signal generated upon touch with a database. APR has been adopted in numerous devices. However, it is not suitable for scenarios where a static object is placed [20]. Knocker [21] is a method of recognizing everyday objects using a single smartphone. This method recognizes a knocked object from the difference in frequency characteristics by recording and analyzing the impulse response generated when the object is knocked with the smartphone containing a microphone. Since this study uses impulse response, recognition is possible in real-time. However, recognition requires the user's knocking action and the object's position cannot be recognized.

Active acoustic sensing

In active acoustic sensing methods, an acoustic signal is applied from a speaker to an object or space, and the response obtained from the microphone is analyzed. For example, attaching microphones and speakers to the human body to propagate acoustic signals is called bio-acoustic sensing, which can be used to detect hand and body posture [22, 23]. Facial expression [23, 24] can also be recognized. These studies show that active acoustic sensing can recognize small changes in muscles and skin. Compared with passive acoustic sensing, active acoustic sensing can even detect contacts and movements that do not generate sounds during the motion, such as soft finger touch. Ono et al. proposed Touch & Activate [25], a system that uses active acoustic sensing to recognize the grasping state of an object with an accuracy of 99.6%. This system makes use of the fact that the frequency response of an object changes when it comes into contact with human skin. Since microphones and speakers are quite commoditized, many researchers in the field of human-computer interaction (HCI) have tried to extend existing devices such as laptops, smartphones, and smartwatches with active acoustic sensing to add operations such as gesture recognition [26–28]. We propose a method that uses this active acoustic sensing principle to turn arbitrary objects into surfaces with object recognition and localization capabilities.

Acoustic-based localization

The localization of an object using acoustic signals from acoustic markers attached to the object is the same as the localization of a sound source at a specific frequency. We employed this method for object localization on a two-dimensional surface. There have been many studies on the methods of source localization.

One promising approach is to employ the TdoA method, based on the time difference of sound arrivals (TdoA) at each microphone. As a method to determine the arrival time difference of sound, PingPongPlus [29] uses the vibration of a ping pong ball touching a ping pong table. Acustico [30] uses the vibration of a surface being touched by users. VersaTouch [31] uses microphones to record the sound propagating from a speaker attached to a finger touching a surface. The microphones determine TdoA by detecting the time when the voltage change of the microphone exceeds a certain level. In addition, Cross-power Spectrum Phase (CSP) analysis is a famous method to estimate the location of sound using two microphones [32]. The CSP method compares signals measured by multiple microphones, calculates the gap between the signals of each microphone using a generalized cross-correlation function, and uses this gap as the TdoA. However, most of the above-mentioned methods are based on the assumption that there is only one sound source; thus, they cannot estimate the location of multiple sources [4]. This is because the acoustic signals interfere with each other when there multiple simultaneous sources.

Surface Acoustic Wave (SAW) is famous acoustic-based localization technology. SAW utilizes the attenuation characteristics of acoustic signals to identify touch locations. This technique is simple in its configuration and can accommodate a variety of objects [20]. However, since SAW measures sound waves traveling on the surface as composite waves, an increase in the types and locations of objects that interfere with these waves can potentially lead to higher learning and computational costs.

Therefore, we attempted to develop a sound source localization system that avoids interference between acoustic signals by separating the frequencies emitted from acoustic markers.

Recognition methodology

This section describes a property of ultrasonic waves and a proposed method for sensing them on the surface of an object based on active acoustic sensing techniques. We also describe the implementation and preliminary experiments of a device that achieves the proposed method.

Sensing principle

The following equation shows the damping of a signal in free space:

$$LOSS = \left(\frac{4\pi r}{\lambda}\right)^2, \tag{1}$$

where r is the distance and λ is the wavelength. The following equation gives the relationship between wavelength and frequency:

$$\lambda = \frac{c}{f}, \tag{2}$$

where c is the speed of sound. Therefore, considering Eq. 1 and Eq. 2, we get the following equation:

$$LOSS = \left(\frac{4\pi rf}{c}\right)^2. \tag{3}$$

Eq. 3 shows that, the higher the frequency of the acoustic signal, the greater the damping due to distance. This is because the acoustic signal loses more energy due to the absorption and attenuation of heat when it propagates due to the stretching effect of the medium.

This paper describes a sound source localization system that utilizes this property. First, a synthetic wave containing low and high-frequency sine waves is emitted from an acoustic marker attached to an object. Second, we measure ultrasonic waves by a microphone on the surface. As the distance between the microphone and the sound source (from the acoustic marker) becomes longer, the higher frequency components of the acoustic signal become more damped. Therefore, as the distance between the microphone and the sound source becomes longer, the difference in the power spectrum between the low and high frequency components of the acoustic signal is expected to increase. This is why we believe that it is possible to estimate the location of an object using the difference between the power spectrum of the low-frequency and high-frequency sound components as a feature value and performing regression analysis using machine learning.

We propose a system that uses active acoustic sensing to recognize object location by attaching sound-emitting markers to obtain and analyze these frequency characteristics. The system analyzes the frequency response as follows: An object's piezoelectric element (acoustic marker) emits a synthetic wave containing multiple frequencies. The piezoelectric element (microphone) attached to the flat plate receives the response on which the object is placed. The system transforms this response into the frequency domain by Fast Fourier Transform (FFT) and analyzes the frequency characteristics.

Note that, as an acoustic signal propagates through an object, the change in the power spectrum at each frequency is affected not only by the distance between the microphone and the source but also by the natural frequency of the object. The natural frequency refers to the frequency at which an object tends to vibrate and, near this frequency, the object resonates and the amplitude increases. Therefore, if the frequency of the acoustic signal emitted from a marker matches this natural frequency of an object or surface, the power spectrum at that frequency will increase regardless of the distance between the microphone and the source. To reduce the influence on the regression analysis of the object's natural frequency or surface, we used a synthetic wave containing many different frequencies. We used artificial waves having frequencies with a total of 36 composite sine waves, from 5 kHz to 40 kHz at the interval of 1 kHz. Since the natural frequency of a surface is a few hundred Hz in width at most, even if one of the frequencies of the used acoustic signal coincides with the natural frequency, our system can estimate object location from the change in the power spectrum of the other frequency bands with distance.

Implementation

Fig. 1 shows the outline of the implemented system. The system consists of an acoustic marker unit made up of a speaker attached to an object, a microphone unit that receives acoustic signals emitted from the marker, and an analysis unit that senses the acoustic signals received by the microphone unit. In the experiments conducted in this paper, all connections and power feeds between amplifiers and acoustic marker, amplifiers and speakers, etc., were wired. For the acoustic marker, we used a bimorph piezoelectric element with a diameter of 21 mm and a thickness of 0.3 mm. We attached the same piezoelectric element to the surface as a microphone using adhesive (Kokuyo's Sticking Insect, Ta-380N). This adhesive can be easily attached again and again, and this system can be applied to any rigid object or surface. When we place an object with an acoustic marker on a surface, acoustic signals propagate through the object to the surface. A TASCAM audio interface (US-16x16) receives the acoustic signals transmitted on the surface with a sampling frequency of 96 kHz. Our system sequentially extracts the data from the piezoelectric response using a Hamming window with a frame size of 512 points and converts it to the frequency domain using FFT. First, the system calculates the power spectrum of the frequency assigned to each object. It identifies which object is placed on the surface based on a predetermined threshold. The spectrums of the frequencies assigned to the objects are then sent to a computer (MacBook Pro, CPU: Intel Core i5 1.4GHz). We use random forest regression models (nTree = 200) to estimate the object locations by regression analysis on the x-axis and y-axis, respectively.

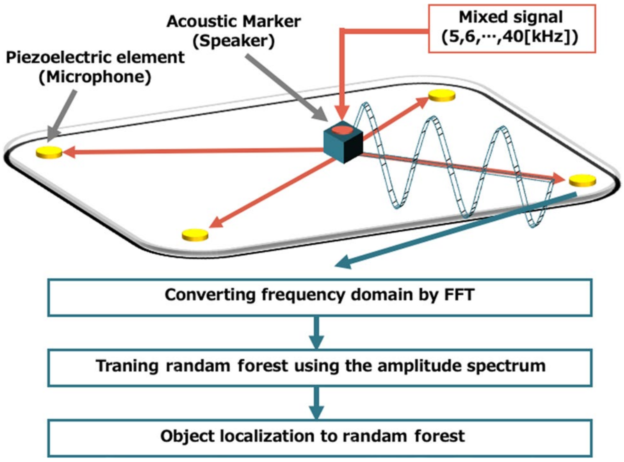
Acoustic Marker (Speaker)
Mixed signal (5,6,⋯,40[kHz])
Piezoelectric element (Microphone)
Converting frequency domain by FFT
Traning randam forest using the amplitude spectrum
Object localization to randam forest

Fig. 1 System overview

Preliminary test

Using our prototype, we evaluated whether the proposed system could estimate the object's location. We placed an object on an acrylic plate (20 cm × 40 cm × 3 mm) and evaluated the estimation accuracy of the object's location. First, we attached an acoustic marker, which emits synthetic waves as described in the previous section, to a wooden block (1 cm × 2 cm × 3 cm). Next, we attached a piezoelectric element to the edge of the surface to serve as a microphone. We placed the object at 10 locations arranged in a straight line at

Table 1 Location estimation results for each algorithm

	R^2	RMSE (cm)	MAE (cm)
Multiple regression	0.895	0.89	0.59
Random forest regression	0.971	0.49	0.17
Support vector regression	0.901	0.67	0.44
Gradient boosting regression	0.961	0.54	0.35

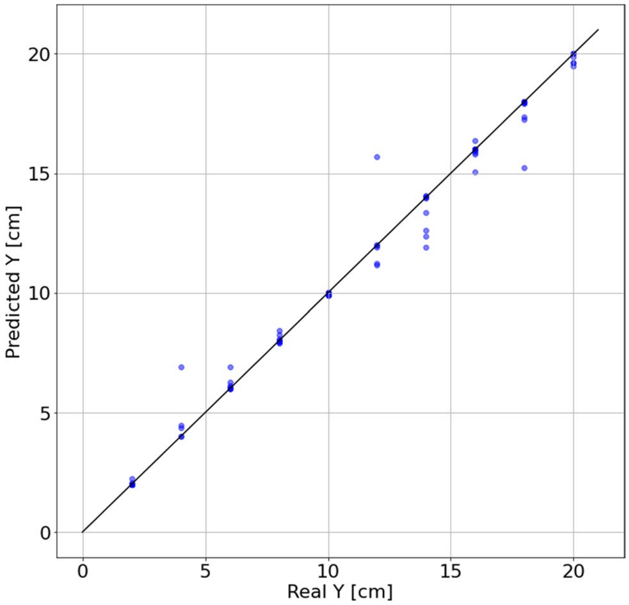

Fig. 2 Actual position and estimated position

2 cm intervals from the microphone and measured 40 times for each. The measurement data were Fourier transformed to obtain the power of the lowest frequency (5k Hz) and the other frequencies (6k, 7k,..., 40k Hz). We used the difference of these 35 spectral strengths as the machine learning features.

We tried four algorithms to create the estimator: Multiple regression analysis, random forest regression, support vector regression (SVR), and gradient boosting regression. We cross-validated the measurement data in five parts and show the results in Table 1.

Here, the closer the value of R^2 is to 1, the better the accuracy, and the smaller the root mean square error (RMSE) and mean absolute error (MAE) are, the closer the actual position and the position estimated by the regression. Looking at the results, the regression using random forest had a good R^2 of 0.971, RMSE of 0.49 cm, and MAE of 0.17 cm. This can be attributed to the fact that random forest works well even with a large number of features. The number of features used in this training was 35. Figure 2 shows the measured values and estimated positions when using random forest. None of the positions deviate significantly from the measured position. We estimated the location with high accuracy even with only one microphone. This indicates that the frequency characteristics obtained by our system have high uniqueness and continuity with the positions of an object and that the proposed system can be used for location estimation.

Multiple objects localization

When recognizing the state of a surface where we placed multiple objects simultaneously, we usually need to acquire training data of all possible combinations of object placement. For example, to distinguish any state where a notebook and/or a pen are placed, we need training data for all states of placing each of the two objects and placing both simultaneously. However, as there are more recognition targets, the number of combinations increases exponentially, thereby making it quite difficult to collect the training data. In this section, we propose a method to independently estimate the location of each object by emitting signals of different frequencies from multiple objects.

When multiple speakers emit sounds in close frequency bands, their sound waves interfere with each other, making position estimation impossible. Therefore, separate frequencies were assigned to each of the multiple objects. For example, suppose that two objects emit sound waves of the following frequencies:

- Object A: from 5 kHz to 40 kHz at an interval of 1 kHz
- Object B: from 5.5 kHz to 40 kHz at an interval of 1 kHz

If Object A and Object B exist on the same surface, the spectrum of the acoustic signal measured by the microphone has 72 peaks of frequencies: 5k, 5.5k, 6k, 6.5k....40k, 40.5k. These spectra obtained by Fourier transform are each divided in frequency bands of every 500 Hz, as shown in Fig. 3, and the spectra of the frequencies assigned to Objects A and B are calculated. After that, we attempted to reduce the training data by estimating the positions of Object A and Object B independently.

Evaluation

Using blocks of cubes, we performed Experiment 1 (E1) to examine the accuracy of position estimation for a single object, Experiment 2 (E2) to evaluate the learning cost, and Experiment 3 (E3) to examine the accuracy of position estimation for multiple objects.

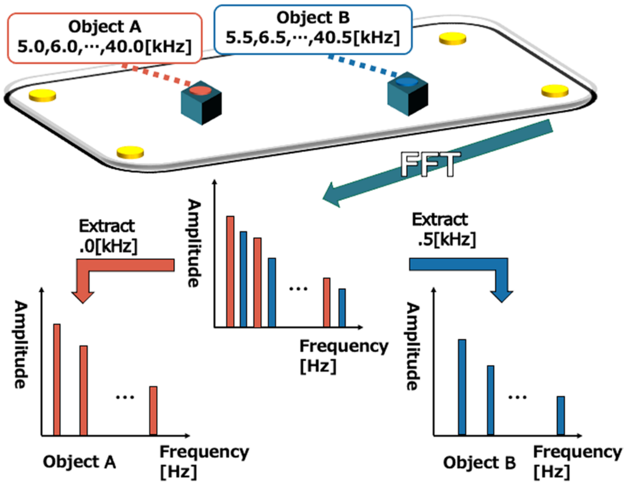

Fig. 3 Split spectrum of a multi-object arrangement

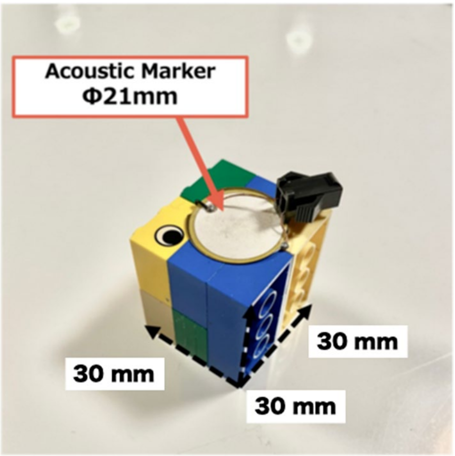

Fig. 4 Acoustic marker and sensing objects used in E1 and E2

E1: single-object localization test

Design and procedure

The purpose of E1 was to test the accuracy of position estimation with a single object placed on our system. We used Lego blocks (30 mm cubes per side) as the target objects to which the markers were to be attached (Fig. 4) and the test area was a 200 mm square with a coordinate system whose origin was the left edge (Fig. 5). We attached piezoelectric elements, which serve as microphones, at a distance of one object (3 cm) from the four corners of the sensing area of the 200 mm square. This prevents the objects from directly contacting the microphones when the objects are placed at the four corners. The test area covered the coordinates from (0,0) to (20,20) and the object placement coordinates were 5 × 5 positions (25 positions in total) spaced 5 cm apart. We placed objects at each position, the system emitted the sound source from the object's marker 40 times each and we obtained the frequency response for each placement. We trained on the x-axis and y-axis for the frequency response data and performed 5-segment cross-validation.

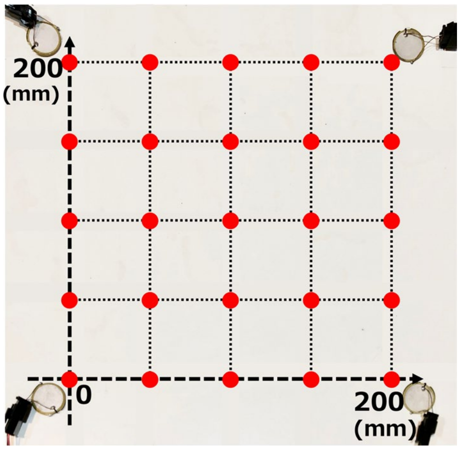

Fig. 5 The sensing area of the 200 mm square for evaluating object localization on 25 points that were 50 mm away from each other

Table 2 Result of single object localization in E1

	R^2	RMSE (cm)	MAE (cm)
x-axis	0.996	0.66	0.29
y-axis	0.996	0.63	0.29

Results and discussion

Table 2 shows the results of the cross-validation. Figure 6 shows the measured and estimated values. The MAE of 0.29 cm in both the x-axis and y-axis directions is equivalent to an error of 0.41 cm ($\fallingdotseq 0.29 \times \sqrt{2}$) in two dimensions. Considering that the object radius used in the experiment is 1.5 cm, this error of 0.41 cm is still acceptable. There was no scatter in error depending on the object's position and we can confirm that the position was estimated with high accuracy at all positions. The size of the object itself may cause this error. We considered the sound source position to be the center of the object, but it is not clear whether the sound propagates from the object's center or not because the sound actually propagates from the acoustic marker to the surface through the object.

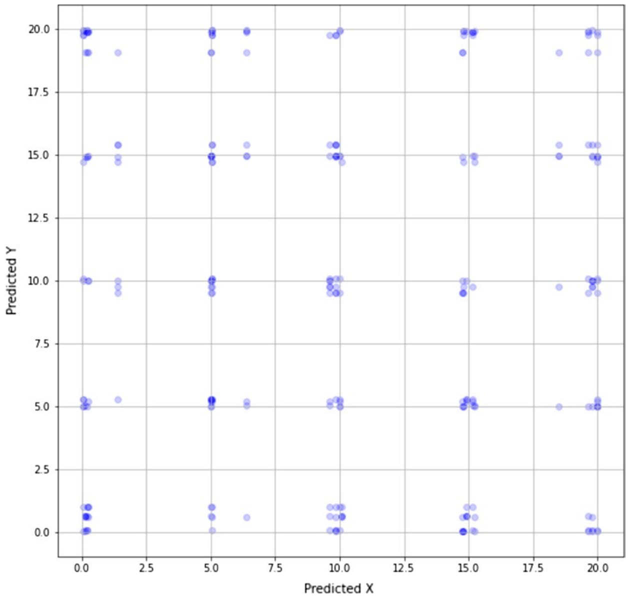

Fig. 6 Estimated location of E1 object

E2: training cost test

Design and procedure

As described in E1, we found that the proposed system using the signal damping difference can estimate the location of a known position on a surface with high accuracy. In this case, it is necessary to prepare training data for each position of the number of objects to be recognized. However, it is not realistic to prepare such a large amount of data for actual use. Therefore, in E2, we tested whether it is possible to estimate not only the known positions but also the unknown positions and evaluated the proposed system by the training cost.

As an experimental procedure, we first attached acoustic markers to Lego blocks (3 cm cubes per side) in the same experimental environment as E1. Next, we placed objects at 21 positions on the diagonal at coordinates (0,0)(1,1)(2,2)...(20,20) and measured 40 times at each position.

After that, we divided the training data and the evaluation data as follows and calculated the error when only some of the position data were used in the training data and the unknown positions were included in the position estimation. For example, in State 6, out of the 21 data points measured, we used data from 4 points ranging from (0,0) to (18,18) in 6 cm increments (Fig. 7).

- State 1: All positions are used for training data. ((0,0)(1,1)...(19,19)(20,20))
- State 2: Each 2 cm position is used as training data. ((0,0)(2,2)...(18,18)(20,20))

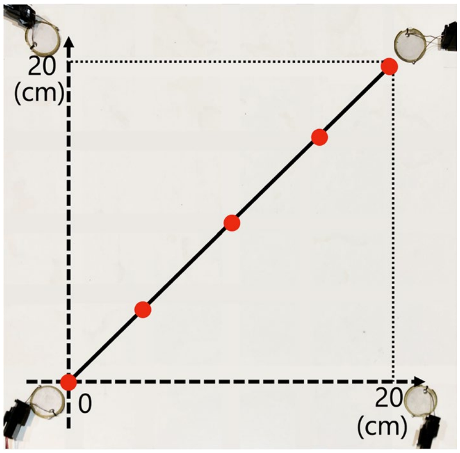

Fig. 7 The sensing area of the 200 mm square for evaluating object localization on 5 points that were 50 mm away from each other (State 5)

- State 3: Each 3 cm position is used as training data. ((0,0)(3,3)...(15,15)(18,18))
- State 4: Each 4 cm position is used as training data. ((0,0)(4,4)...(16,16)(20,20))
- State 5: Each 5 cm position is used as training data. ((0,0)(5,5)...(15,15)(20,20))
- State 6: Each 6 cm position is used as training data. ((0,0)(6,6)(12,12)(18,18))

Results and discussion

Table 3 and Fig. 8show the experimental results. Here, MAE indicates the error between the actual and estimated positions. From Fig. 8, we found that the estimation error increases as the number of positions used for training decreases. This can be attributed to reducing the number of data used for training. However, in State 3, the error was larger than in States 4 and 5, where there was fewer training data. This may be because the last position of the placement data used for the training data in States 4 and 5 was (20,20), while the last position in State 3 was (18,18). When creating a regression model for position estimation using machine learning, it is more likely that a good regression model will

Table 3 Result of location detection in E2

	R^2	RMSE (cm)	MAE (cm)
State 1	0.982	0.59	0.24
State 2	0.970	1.04	0.58
State 3	0.805	2.67	1.42
State 4	0.935	1.53	0.98
State 5	0.851	2.33	1.61
State 6	0.728	3.15	1.98

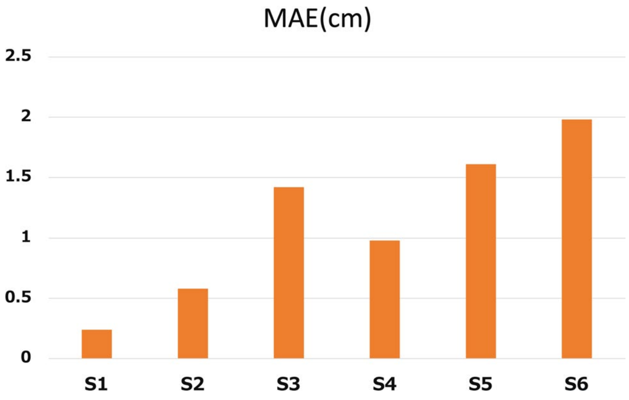

Fig. 8 Result of MAE with decreasing training cost

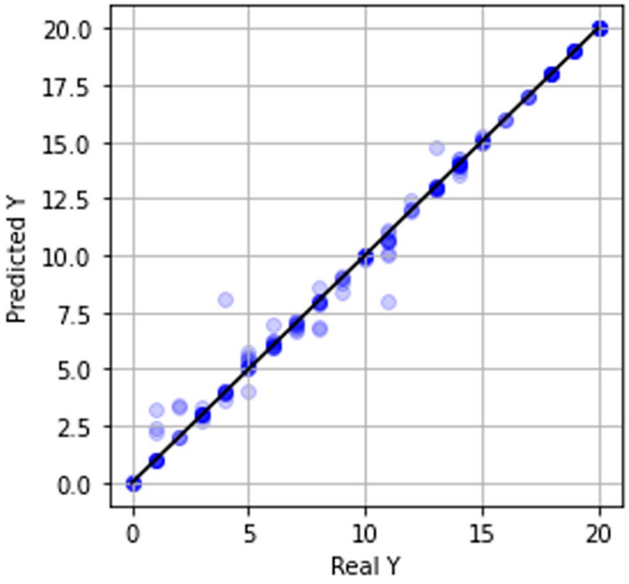

Fig. 9 Result of State 1

be made if there is an end and a beginning of continuous data. Therefore, the regression models created in States 4 and 5 may have been more accurate than the regression models made in State 3. This indicates that, for training, it is better to place objects on the surface at each of its sensing range vertices in our system.

The Figs. 9, 10, 11 shows that the wider the interval between training data, the more the estimation results were stair-stepped, i.e., the output was pulled to the nearest training point rather than pure regression. Our results suggested that there is some correlation between sound wave attenuation and object location. Elucidating this property may allow for a more linear regression model.

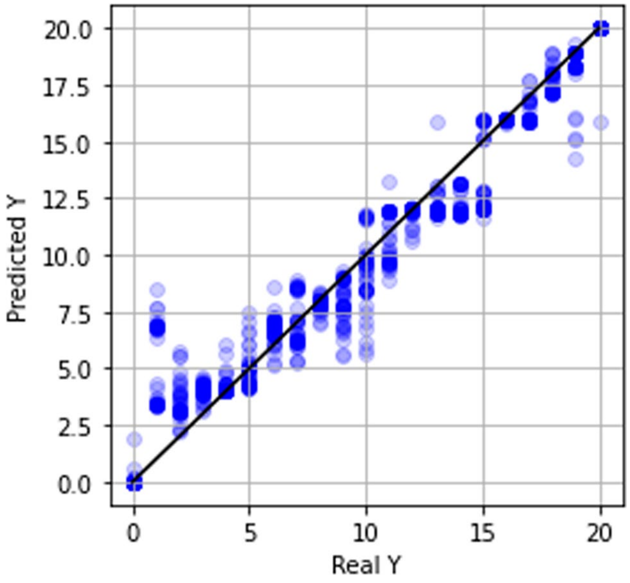

Fig. 10 Result of State 4

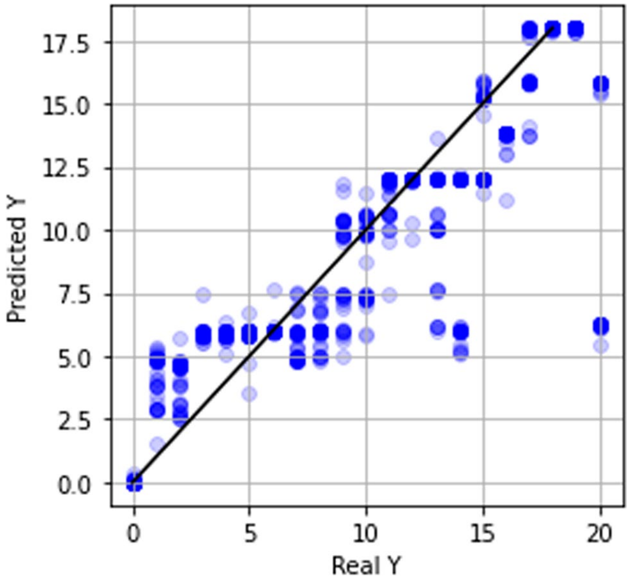

Fig. 11 Result of State 6

The MAE in State 4 was 0.98 cm. This indicates that the system is able to estimate object position on the diagonal of a 200 mm square with an error of 0.98 cm by placing objects in only 5 positions. This estimation error is sufficiently small considering that one side of the object is 3.0 cm. This also suggests that the frequency response obtained by our system has a high continuity and linearity concerning the object's position on the surface. Therefore, it is unnecessary to learn all the locations to be recognized and the training cost can be reduced.

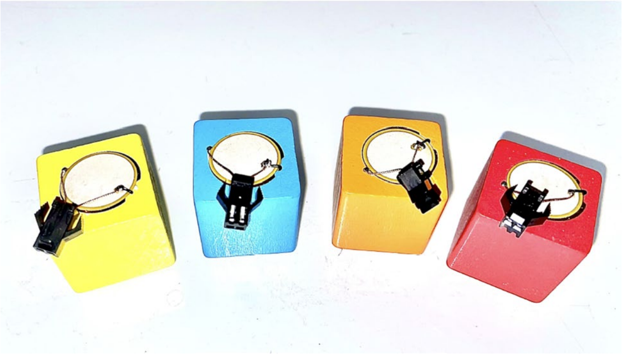

Fig. 12 2.0 cm wooden cubic blocks per side with acoustic marker

E3: multiple objects recognition test

Design and procedure

In E3, we evaluated the accuracy of our method for multiple object placement by testing it on actual objects on a surface. This experiment first attached acoustic markers to four 2.0 cm wooden cubic blocks per side (Fig. 12) in the same experimental environment as E1 and E2. We placed the objects at five positions (0,0), (5,5), (10,10), (15,15), and (20,20) in the diagonal direction of the surface plane. The estimated number of objects was set to 2–4. The frequency assignment of acoustic markers for each object number was set as follows.

Two objects (Each 500 Hz)

- Object A: 5k, 6k,..., 39k, 40k Hz
- Object B: 5.5k, 6.5k,..., 39.5k, 40.5k Hz

Three objects (Each 333 Hz)

- Object A: 5k, 6k,..., 39k, 40k Hz
- Object B: 5.333k, 6.333k,..., 39.333k, 40.333k Hz
- Object C: 5.666k, 6.666k,..., 39.666k, 40.666k Hz

Four objects (Each 250 Hz)

- Object A: 5k, 6k,..., 39k, 40k Hz
- Object B: 5.25k, 6.25k,..., 39.25k, 40.25k Hz
- Object C: 5.5k, 6.5k,..., 39.5k, 40.5k Hz
- Object D: 5.75k, 6.75k,..., 39.75k, 40.75k Hz

We next describe the experimental procedure. In the case of two objects, we first placed only Object A at the five positions, (0,0), (5,5), (10,10), (15,15), and (20,20), and measured 40 times at each position. In the same way, we placed only Object B in the same five locations as before and measured 40 times at each location. Next, we placed Object A and Object B on the surface at the five positions, (0,0), (5,5), (10,10), (15,15), and (20,20), simultaneously and measured 40 times at each position. The placement pattern of the two objects is ${}_{20}C_2 = 190$. For all the measurement data obtained in this way, we estimated the position of Object A and Object B for the pattern where both were placed simultaneously and calculated the error using the data where each Object A and Object B was placed on the surface for training. In the case of three objects and four objects, we estimated the position of each object from the data of a single object in the patterns where three or four objects were placed and calculated the estimation error.

Results and discussion

Fig. 13 shows the MAE for different numbers of objects, together with the mean absolute error for a single object as described in Sec. 3.3. The mean error with one object was 0.29 cm; with two objects, the mean error was 0.84 cm; with three objects, the mean error was 1.67 cm; and with four objects, the mean error was 1.83 cm. These results show that, the more objects there are on the surface, the bigger the error becomes. This may be because the frequency of each object becomes greater as the number of objects increases and the acoustic signals interfere with each other. In this study, the acoustic signals emitted from the marker were synthesized from 36 sine waves in 1 kHz segments. In the future, this error can be reduced by reducing the number of sine waves and widening the difference in frequency between each object. However, if the number of synthesized sine waves is reduced, the number of features used for learning is also reduced, which may increase the estimation error. Another possible cause of the error is that the acoustic signal propagating from one object diffracts off the other object, changing the propagation distance until it reaches the microphone and affecting the estimated position.

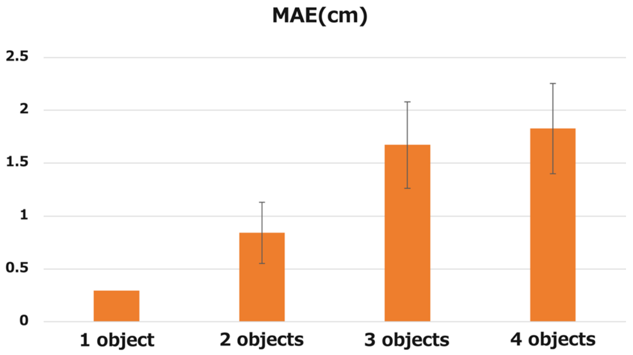

Fig. 13 Object localization accuracy evaluated on 1–4 different objects

The results obtained by E3 show that the mean error between each object's estimated and measured position was 1.82 cm even when four objects are placed on the surface. This is smaller than that of a related study [33]. Through this experiment, we show that it is possible to estimate the locations of multiple objects using the training data of a single object. In addition, the training data required to recognize four object placements for five positions using machine learning is generally 120 data (= $_5P_4$). In contrast, in our system, $4(numberofobjects) \times 5(locations) = 20$ data. Therefore, our system saves 83.3% of the training data collection effort.

Discussion

Potential applications

The E1 and E2 results show our system is capable of estimating 21 object positions with an error of 0.98 cm by placing objects at only 5 positions and training them. Furthermore, the E3 results suggest that our system recognizes the placement of four objects with an accuracy of 1.83 cm, saving 83.3% of the effort needed to collect the training data. While further exploration would be needed to improve accuracy, these findings adequately demonstrate the potential as a context-aware sensing device that can be used for applications as follows:

First, the system will easily make everyday objects tangible by attaching piezoelectric elements to objects and surfaces. For example, we attach a marker to an existing figure and let the PC recognize the type and position of the figure and combine it with artificial reality (AR) technology to create a tangible toy for interaction. In general, such toys can only be applied to objects with built-in sensors, but our system has no such limitations. It can easily, tangibly use everyday objects that we are familiar with. What is required then is that the part of the object that touches the surface be of a material that propagates sound. As we did in this experiment, this can be realized by embedding a wireless earphone-like mechanism somewhere on or inside a object.

Second, our system can be used without a marker by attaching to any device that has a speaker function. For example, by emitting acoustic signals from a smartphone and sensing them with our system, the position of the smartphone can be continuously acquired14. Therefore, this system can extend the interaction of existing acoustic devices, such as operating a PC using a smartphone as a controller.

Limitation and future work

The first limitation of our system is that the acoustic signals emitted from the marker are in the audible range. Therefore, we need to improve the hardware, such as wrapping the marker itself with soundproof material to prevent the acoustic signal from leaking into the air or changing the acoustic signal so that it is not in the audible range. Generally, the human audible range is between 20 Hz and 20 k Hz. Therefore, this problem can be solved by using an acoustic signal synthesized from a low-frequency sine wave (below 20 Hz) and a high-frequency sine wave (above 20k Hz). However, better hardware such as contact microphones with higher sensitivity is needed to simultaneously sense lower and higher frequencies.

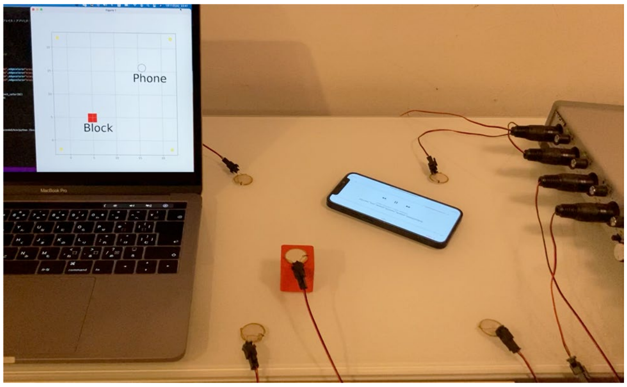

Fig. 14 Application examples: Recognizing the location of a smartphone

The second limitation is the Doppler shift when we move the object quickly. Our system recognizes the frequency of the acoustic signal emitted from the marker and identifies the object. Therefore, the Doppler shift that occurs when moving objects may cause a change in the frequency, leading to misrecognition of the object. In the future, we need to study the change in frequency characteristics caused by the object's movement and create a mathematical model to calculate the frequency change caused by the Doppler shift.

The third limitation is the validity of this synthesized wave. Our system uses a synthesized sine wave as the acoustic signal emitted from the marker, but we need to examine it. Therefore, we will change the number of synthesized sine waves and their frequency range, investigate the effect on the estimation error, and compare it with the error when placing multiple objects.

The fourth limitation, all moving objects must emit sound. If an object without a speaker is placed on the surface, the acoustic properties on the surface will change depending on the object, and the machine learning will need to be redone. In our previous research [34], a speaker and microphone were mounted on a similar surface, and several objects without speakers were placed on the surface. By calculating the attenuation of sound waves traveling over the surface, we were able to identify the type and location of the multiple objects placed, but there was no law for the effect of moving object location on the acoustic properties. Therefore, it is difficult to apply the present method in situations where objects without speakers are randomly placed and removed.

Finally, because the proposed method must emit sound, a battery is required for wireless operation. It cannot be driven indefinitely, and the battery life would be comparable to that of wireless earphones.

Conclusion

In this paper, we proposed an acoustic marker system that uses active acoustic sensing to recognize the placement of objects. The prototype consists of an audio interface and piezoelectric elements easily and tangibly attached to objects and surfaces. Using the prototype, we confirmed that the system recognizes the placement of a single object with a mean absolute error of 0.41 cm. We also verified the training cost of our system and found that 21 positions could be estimated with an error of 0.98 cm by training only five positions. We then tested the recognition performance of our system when placing multiple objects simultaneously and evaluated it. As a result, we found that each object could be estimated with a mean error of 1.82 cm even when four objects were placed on the surface simultaneously. This experiment shows that our system can recognize multiple objects simultaneously from single-object training data.

In the future, we will improve the hardware by wrapping the marker with soundproof material to prevent acoustic signals from leaking into the air. We will then investigate the validity of the acoustic signals emitted from the markers by changing the number of synthesized sine waves and their frequency range. In this way, we can reduce the error further by optimizing the acoustic signal used.

Acknowledgements This work was partially supported by JSPS KAKENHI Grant Number JP20H04228.

Declarations

Conflict of interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Villar N, Cletheroe D, Saul G et al (2018) Project zanzibar: a portable and flexible tangible interaction platform. In: Proceedings of the 2018 CHI conference on human factors in computing systems. Association for computing machinery, New York, NY, USA, CHI'18, pp 1–13. https://doi.org/10.1145/3173574.3174089
2. Redmon J, Farhadi A (2018) Yolov3: an incremental improvement. CoRR arxiv:1804.02767
3. Bränzel A, Holz C, Hoffmann D et al (2013) Gravityspace: tracking users and their poses in a smart room using a pressure-sensing floor. In: Proceedings of the SIGCHI conference on human factors in computing systems. Association for computing machinery, New York, CHI '13, pp 725–734 https://doi.org/10.1145/2470654.2470757
4. Chan YT, Ho K (1994) A simple and efficient estimator for hyperbolic location. IEEE Trans Signal Process 42(8):1905–1915
5. Dietz PH, Eidelson BD (2009) Surfaceware: dynamic tagging for microsoft surface. In: Proceedings of the 3rd international conference on tangible and embedded interaction. Association for computing machinery, New York, NY, USA, TEI'09, pp 249–254. https://doi.org/10.1145/1517664.1517717
6. Weiss M, Wagner J, Jansen Y, et al (2009) Slap widgets: Bridging the gap between virtual and physical controls on tabletops. In: Proceedings of the SIGCHI conference on human factors in computing systems. Association for computing machinery, New York, NY, USA, CHI'09, pp 481–490. https://doi.org/10.1145/1518701.1518779
7. Zhang Y, Yang CJ, Hudson SE et al (2018) Wall++: room-scale interactive and context-aware sensing. In: Proceedings of the 2018 CHI conference on human factors in computing systems, New York, NY, USA, CHI'18, https://doi.org/10.1145/3173574.3173847
8. Gong J, Wu Y, Yan L et al (2019) Tessutivo: contextual interactions on interactive fabrics with inductive sensing. In: Proceedings of the 32nd annual ACM symposium on user interface software and technology. association for computing machinery, New York, NY, USA, UIST'19, pp 29–41. https://doi.org/10.1145/3332165.3347897
9. Bränzel A, Holz C, Hoffmann D et al (2013) Gravityspace: tracking users and their poses in a smart room using a pressure-sensing floor. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, NY, USA, CHI'13, pp 725–734. https://doi.org/10.1145/2470654.2470757
10. Harrison C, Tan D, Morris D (2011) Skinput: appropriating the skin as an interactive canvas. Commun ACM 54(8):111–118. https://doi.org/10.1145/1978542.1978564
11. Paradiso JA, Leo CK, Checka N et al (2002) Passive acoustic knock tracking for interactive windows. In: CHI '02 extended abstracts on human factors in computing systems. ACM, New York, NY, USA, CHI EA'02, pp 732–733. https://doi.org/10.1145/506443.506570
12. Harrison C, Hudson SE (2008) Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. In: Proceedings of the 21st annual ACM symposium on user interface software and technology. Association for Computing Machinery, New York, NY, USA, UIST '08, pp 205–208, https://doi.org/10.1145/1449715.1449747
13. Chen M, Yang P, Xiong J et al (2019) Your table can be an input panel: acoustic-based device-free interaction recognition. Proc ACM Interact Mob Wearable Ubiquitous Technol 3(1):3:1-3:21. https://doi.org/10.1145/3314390
14. Harrison C, Schwarz J, Hudson SE (2011) Tapsense: enhancing finger interaction on touch surfaces. In: Proceedings of the 24th annual ACM symposium on user interface software and technology. ACM, New York, NY, USA, UIST'11, pp 627–636 https://doi.org/10.1145/2047196.2047279
15. Lopes P, Jota R, Jorge JA (2011) Augmenting touch interaction through acoustic sensing. In: Proceedings of the ACM international conference on interactive tabletops and surfaces. ACM, New York, NY, USA, ITS'11, pp 53–56 https://doi.org/10.1145/2076354.2076364
16. Sim JM, Lee Y, Kwon O (2015) Acoustic sensor based recognition of human activity in everyday life for smart home services. Int J Distrib Sen Netw. https://doi.org/10.1155/2015/679123
17. Yatani K, Truong KN (2012) Bodyscope: a wearable acoustic sensor for activity recognition. In: Proceedings of the 2012 ACM

conference on ubiquitous computing. ACM, New York, NY, USA, UbiComp'12, pp 341–350 https://doi.org/10.1145/2370216.2370269

18. Amento B, Hill W, Terveen L (2002) The sound of one hand: A wrist-mounted bio-acoustic fingertip gesture interface. In: CHI'02 extended abstracts on human factors in computing systems. ACM, New York, NY, USA, CHI EA'02, pp 724–725. https://doi.org/10.1145/506443.506566
19. Wu J, Harrison C, Bigham JP et al (2020) Automated class discovery and one-shot interactions for acoustic activity recognition. In: Proceedings of the 2020 CHI conference on human factors in computing systems. Association for Computing Machinery, New York, NY, USA, CHI '20, pp 1–14. https://doi.org/10.1145/3313831.3376875
20. Gao S, Yan S, Zhao H et al (2021) Touch detection technologies. In: Touch-based human-machine interaction: principles and applications. Springer, pp 19–89. https://doi.org/10.1007/978-3-030-68948-3_3
21. Gong T, Cho H, Lee B et al (2019) Knocker: vibroacoustic-based object recognition with smartphones. Proc ACM Interact Mob Wearable Ubiquitous Technol. https://doi.org/10.1145/3351240
22. McIntosh J, Marzo A, Fraser M et al (2017) Echoflex: hand gesture recognition using ultrasound imaging. In: Proceedings of the 2017 CHI conference on human factors in computing systems. Association for Computing Machinery, New York, NY, USA, CHI'17, pp 1923–1934. https://doi.org/10.1145/3025453.3025807
23. Iravantchi Y, Zhang Y, Bernitsas E et al (2019) Interferi: gesture sensing using on-body acoustic interferometry. In: Proceedings of the 2019 CHI conference on human factors in computing systems. Association for Computing Machinery, New York, NY, USA, CHI'19, pp 1–13 https://doi.org/10.1145/3290605.3300506
24. Amesaka T, Watanabe H, Sugimoto M (2019) Facial expression recognition using ear canal transfer function. In: Proceedings of the 23rd international symposium on wearable computers. Association for Computing Machinery, New York, NY, USA, ISWC'19, pp 1–9 https://doi.org/10.1145/3341163.3347747
25. Ono M, Shizuki B, Tanaka J (2013) Touch and activate: adding interactivity to existing objects using active acoustic sensing. In: Proceedings of the 26th annual ACM symposium on user interface software and technology. ACM, New York, NY, USA, UIST'13, pp 31–40 https://doi.org/10.1145/2501988.2501989
26. Gupta S, Morris D, Patel S et al (2012) Soundwave: using the doppler effect to sense gestures. In: Proceedings of the SIGCHI conference on human factors in computing systems. Association for Computing Machinery, New York, NY, USA, CHI'12, pp 1911–1914. https://doi.org/10.1145/2207676.2208331
27. Laput G, Harrison C (2019) Sensing fine-grained hand activity with smartwatches. In: Proceedings of the 2019 CHI conference on human factors in computing systems. Association for Computing Machinery, New York, NY, USA, CHI'19, pp 1–13 https://doi.org/10.1145/3290605.3300568
28. Nandakumar R, Iyer V, Tan D et al (2016) Fingerio: using active sonar for fine-grained finger tracking. In: Proceedings of the 2016 CHI conference on human factors in computing systems. Association for Computing Machinery, New York, NY, USA, CHI'16, pp 1515–1525. https://doi.org/10.1145/2858036.2858580
29. Ishii H, Wisneski C, Orbanes J et al (1999) Pingpongplus: Design of an athletic-tangible interface for computer-supported cooperative play. In: Proceedings of the SIGCHI conference on human factors in computing systems. Association for Computing Machinery, New York, NY, USA, CHI'99, pp 394–401 https://doi.org/10.1145/302979.303115
30. Gong J, Gupta A, Benko H (2020) Acustico: Surface tap detection and localization using wrist-based acoustic TDOA sensing. In: Proceedings of the 33rd annual ACM symposium on user interface software and technology. Association for Computing Machinery, New York, NY, USA, UIST'20, pp 406–419 https://doi.org/10.1145/3379337.3415901
31. Shi Y, Zhang H, Cao J et al (2020) Versatouch: a versatile plug-and-play system that enables touch interactions on everyday passive surfaces. In: Proceedings of the augmented humans international conference. Association for Computing Machinery, New York, NY, USA, AHs'20 https://doi.org/10.1145/3384657.3384778
32. Knapp C, Carter G (1976) The generalized correlation method for estimation of time delay. IEEE Trans Acoust Speech Signal Process 24(4):320–327. https://doi.org/10.1109/TASSP.1976.1162830
33. Kreskowski A, Wagner J, Bossert J et al (2015) Mobat: sound-based localization of multiple mobile devices on everyday surfaces. In: Proceedings of the 2015 international conference on interactive tabletops and Surfaces. Association for Computing Machinery, New York, NY, USA, ITS'15, pp 247–252. https://doi.org/10.1145/2817721.2823488
34. Iwase D, Itoh Y, Shin H et al (2019) Sensesurface: using active acoustic sensing to detect what is where (in Japanese). Journal of Information Processing (JIP) 60(10):1869–1880

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.