IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 14 November 2023, accepted 6 December 2023, date of publication 15 December 2023, date of current version 20 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3343500

RESEARCH ARTICLE

Smatable: A Vibration-Based Sensing Method for Making Ordinary Tables Touch-Interfaces

MAKOTO YOSHIDA[1], TOMOKAZU MATSUI[1], (Member, IEEE), TOKIMUNE ISHIYAMA[1], MANATO FUJIMOTO[2,3], (Member, IEEE), HIROHIKO SUWA[1,3], (Member, IEEE), AND KEIICHI YASUMOTO[1,3], (Member, IEEE)

[1]Graduate School of Science and Technology, Nara Institute of Science and Technology, Nara 630-0192, Japan
[2]Graduate School of Informatics, Osaka Metropolitan University, Osaka 558-8585, Japan
[3]RIKEN Center for Advanced Intelligence Project AIP, Tokyo 103-0027, Japan

Corresponding author: Makoto Yoshida (yoshida.makoto.yn3@is.naist.jp)

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP20H04177 and Grant JP19H05665.

ABSTRACT In recent years, the equipment that makes up smart homes is required not only to be functional, but also to be integrated with the design and aesthetics of the living space. Among them, interfaces that directly touch the human eye and hands are the key to maintaining design, but there were many issues in terms of integration with design and aesthetics of living spaces. In this paper, we propose an interface system that operates existing furniture by touching it as a new interface that integrates beautifully into the living space. The proposed system detects user operations with four small vibration sensors attached to hidden locations of existing furniture and uses deep learning to learn the vibrations when a person touches the furniture. Using this method, thick materials difficult to achieve with normal capacitive touch sensors can be utilized. In the experiment, a dining table was used as a representative piece of furniture, and the accuracy of detecting the direction in which three participants swiped in four directions on the table was verified. As a result of the experiment, the accuracy was confirmed by Leave-One-Person-Out-Cross-Validation using 3 sessions of swipe data for each individual for 3 participants, and the accuracy was 0.67. Furthermore, we verified the accuracy of a trained model created by adding only one session of evaluation target data to each learning dataset used in the Leave-One-Person-Out-Cross-Validation. As a result, the accuracy reached 0.90, achieving practical precision.

INDEX TERMS Touch interface, operation recognition, vibration sensor, deep learning.

I. INTRODUCTION

As housing equipment becomes smarter, smart homes are becoming more familiar, and various studies are being conducted not only on smart home appliances but also on smart furniture [1], [2], [3]. Additionally, smart furniture is being researched to provide comfort for the elderly [4]. Furthermore, in recent years, smart home appliances and housing equipment are required not only to be functional but also to have design features such as beauty and texture that harmonize with the living space without making people aware of their presence. The interface, in particular, directly links the dweller and the equipment and is the key to harmony among people, equipment, furniture, and living space design. Existing input interfaces currently used in smart homes include remote controls and smartphone applications. However, various studies have been conducted to realize more human-friendly interfaces.

For example, input methods using cameras, infrared rays, or a combination of both have been studied so that a person can operate a device such as a remote control without holding it using the person's movements, gestures, and so on [5], [6], [7], [8], [9], [10], [11]. Also, to combine display and sensor devices for more comfortable operation, research is being conducted on interfaces that utilize both a highly expressive Graphical User Interface (GUI) using projection

The associate editor coordinating the review of this manuscript and approving it for publication was Nagendra Prasad Pathak.

mapping and information on human movement captured by cameras [12], [13], [14]. These methods have the advantage of being operable without the need to hold the device and can be combined with GUI to support hierarchical operations such as menu screens. However, cameras and motion sensors must be positioned where the sensors can see the target object. Similarly, projection mapping must be positioned where people are not in shadow. Thus, cameras, motion sensors, and projection mapping have occlusion problems and restrict the furniture arrangement in the home, which in turn imposes restrictions on the living space design. Moreover, the visibility of sensors to users poses a challenge. Even if an intuitive interface is achieved, maintaining the aesthetic harmony and texture of furniture within the space due to sensor placement remains a challenge (Issue 1). In addition to this, privacy concerns exist when cameras and other image recognition devices are placed indoors (Issue 2).

On the other hand, interfaces using voice recognition [15] are also being used. Interfaces using voice recognition have the advantage of saving installation space and being easy to introduce. However, there are issues such as the need to generate voice several times by the adjustment amount when continuously changing the volume adjustment, for example, which makes the operation cumbersome and unintuitive (Issue 3). Additionally, as a method different from voice recognition, some methods use a microphone to determine the movement of a person's operation from an acoustic signal [16], [17]. Some research extends the input of more advanced characters using acoustic signals [18], [19]. However, like cameras, privacy concerns exist when microphones are installed indoors as smart home controllers. Therefore, as a sensor device that recognizes human movement without concern about privacy, a sensor device using electromagnetic waves is expected, and research is being conducted. However, the method using electromagnetic waves has the problem of lowering the recognition accuracy due to changes in the radio wave environment caused by the metal furniture or changes in the layout of the room [20], [21], [22], [23], [24]. This imposes restrictions not only on the design of living spaces but also on the materials used for furniture, leaving issues with the texture, aesthetics, and design of the materials that people see and touch (Issue 4).

Meanwhile, in the display interface, an important interface as well as the input device, research is being conducted on a hidden interface that emphasizes design and aesthetics and can be blended into the user's living space. Olwal et al. [25] proposed a display interface in which display devices can be embedded under panels, such as the panels of furniture or decorative panels of home appliance bodies, and made invisible by turning off the display devices when not in use. Hidden interfaces are particularly important in the home where appearance and design are important, as they can be made invisible or completely unnoticeable except when the interface is used, and can also use the texture of everyday furniture and building materials as they are. However, since the above study mainly discusses display devices, the conventional electrostatic method is used for the input interface. In general, capacitive touch sensors have extremely low sensitivity under thick materials, and there are issues, such as the difficulty of using capacitive touch sensors on furniture made of metal materials. As such, there is a problem in concealing and installing it while maintaining the texture of various materials used in existing furniture.

Based on the discussion so far, for the interface to blend seamlessly into the user's living space without compromising the design and aesthetics of the room, the interface should be invisible to humans. In addition, it is essential to allow flexibility in the design of the space where the interface is used, the furniture's material, and consider the privacy issue. To address these requirements, issues 1-4 must be resolved. However, concerning input interfaces in particular, there remains a lack of discussion regarding input devices capable of retaining the materials and textures of devices and furniture.

In this paper, we propose an input interface that can be operated by touching the existing furniture and solving the above issues 1 to 4. We design and develop the input interface to recognize input operations using only the vibrations generated when a user swipes the furniture by attaching four vibration sensors using thin piezoelectric devices to hidden places on furniture and equipment. Two vibration sensors are arranged horizontally for the X-axis, and the other two are arranged vertically for the Y-axis. Hereafter, we call an output of the sensors as channels or CHs. The system detects the vibrations generated when a person touches the furniture with a finger using sensors on each of the four CHs and classifies the direction in which the person swipes the furniture into four different types by learning the signal data obtained from the sensors using deep learning. These sensors can be used not only by attaching them to existing furniture but also by embedding them inside furniture or behind thick wall panels, making them completely invisible to the user. In addition, unlike electrostatic touch sensors, this method is less affected by the thickness of the material. Therefore, the sensor can be placed behind thicker materials, allowing the operation surface to use the material's texture, such as thick wood. Additionally, it can handle metallic materials, which is difficult to achieve with electrostatic touch sensors. This capability enables the realization of an input interface that complements the design of indoor furniture and building components made from textured materials without compromising their aesthetics.

In this study, we chose a table as one of the furniture to install the interface and studied the recognition of the 4-direction swipe operation (UP, DOWN, RIGHT, LEFT) by tracing the tabletop with a finger. By devising the preprocessing in the learning process of data obtained from sensors and the network structure of deep learning, we were able to improve the accuracy by more than 15% compared to the previous study [26]. This improvement in accuracy has made the system worthy of practical use, although it requires some processing ingenuity, such as retrying in case of false detection. Specifically, the following two points were improved:

- We carefully observed the changes in the fundamentals and harmonics of the vibration signal during swiping, and considering that the changes in the harmonic content are important for swipe direction detection, we modified the Short Time Fourier Transform (STFT) parameters to increase the frequency resolution.
- To learn more detailed features during learning, we increased the number of layers of the deep learning network (from 5 layers to 7 layers) and introduced a residual connection.

As a result of the accuracy improvement, in Leave- One-Person-Out-Cross-Validation (LOPOCV), The accuracy was improved by approximately 15% when judging the swipe direction of the participant to be estimated, achieving an accuracy of 0.67. Furthermore, in addition to the above conditions, the accuracy of the model trained by adding the swipe data of an arbitrary estimation target participant to the learning data for only one session improved the accuracy by approximately 16%, and the accuracy reached 0.90. The main contributions of this paper are summarized as follows:

- First, we have shown that the improvement of the resolution of the frequency axis and the improvement of the Convolutional Neural Network (CNN) network lead to the improvement of the accuracy, and have developed a new classification algorithm with higher accuracy than the conventional method.
- Second, we confirmed that a practical level of accuracy can be achieved by registering data in advance by users who use the interface. As a result of this improved accuracy, it is possible to create a furniture interface that allows users to swipe their fingers across the furniture's surface by placing the sensor in an invisible location on the furniture and using vibration while keeping the same material.

The remainder of this paper is organized as follows. Section II introduces related work on in-home interfaces and discusses its challenges. Section III describes the requirements that an interface for home use should aim for, as well as the hardware and algorithms of the proposed swipe direction detection system. Section IV evaluates our proposed method. Section V discusses touch sensing using vibration and improvements. Finally, we conclude our paper in Section VI.

II. RELATED WORK

In this section, we introduce some existing studies related to our research, summarize their problems, and then clarify the research goals of this paper.

A. INPUT INTERFACE USING CAMERA

In recent years, research has been conducted on camera-based gesture recognition as an interface. Goto et al. [12] used a pan-tilt camera and projector to project a GUI image on a table or wall in the home and used video footage of user actions on the projected image to detect hand gestures as if it were a touch screen. Simone et al. [13] have constructed an interface by combining a camera, a projector, and an infrared laser to realize the recognition of human actions, including multi-touch, on the projected image.

However, not only are these methods large structures, but they also present occlusion challenges, and the camera must be installed so that the projection surface and hands are within the camera's angle of view yet unobstructed by obstructions. As a result, there is a challenge of restricting the positioning relationship of the furniture. Also, apart from privacy concerns when the camera is installed in a room, there are also psychological considerations when the camera is not in use. This is because the camera is positioned where it remains clearly visible to the user.

Moreover, considering practical aspects, numerous situations arise where the home lighting is dimmed, causing a reduction in the light captured by the image sensor. This brightness decrease makes it challenging to increase the frame rate, leading to unclear images and greater difficulty in identifying moving objects. Additionally, operation-wise, there is the issue of dealing with substantial information in the image data, which demands computational expenses for preprocessing.

B. INPUT INTERFACE USING A MICROPHONE

Braun et al. [27] used multiple contact microphones on a desk to detect events such as knocks and swipes combined with machine learning. By using multiple microphones, it is expandable so that it can be used as a backup in case of microphone failure, signal synthesis, and external noise judgment. This study successfully categorizes impact events involving object hits and swiping events. Nevertheless, a limitation lies in the failure to detect the direction of the swipe, posing a challenge to achieving intuitive operation.

Mayank et al. [28] proposed a system for association and information transfer between multiple devices, such as smartphones, placed on a mutually shared plane. The system recognizes various gestures such as swiping, pinching, and expanding with fingers on a plane shared by multiple smartphones, using 72 features and the k-nearest neighbor method, using sensors and contact microphones built into the smartphones. Furthermore, it is also possible to detect the direction of the swipe. However, since the surface to be operated is limited to the plane on which the device is placed, it is insufficient to be used as a hidden interface system that is not visible to humans.

C. CAPACITIVE TOUCH SENSOR

Capacitive touch sensors have become widely used as touch interfaces in smartphones, etc., and various research is still being conducted to improve convenience further. Pourjafarian et al. [29] proposed a capacitive touch sensor system that runs on a general-purpose microcontroller and requires no specialized hardware. Parilusyan et al. [30] also aim to integrate sensors into physical materials and are developing a modular hardware platform that allows a variety of materials to be used. However, these methods require sensors to be spread over the entire operation surface, and because they use a capacitive method, they have not been verified when using metal materials or when embedded under thick materials. For this reason, there are challenges when sensing while taking advantage of existing furniture designs. Wu et al. [31] have developed unique touch sensors that can identify the front and back sides of the touch surface, respectively. These studies are very good in that soft materials such as clothing can be used as touch sensors, but because this method is also a capacitance-based sensor, it is difficult to place the sensor under thick boards such as furniture.

D. FORCE SENSOR

Lee et al. [32] are developing a physiological signal monitoring bed for infants with a load cell sensor, a type of force sensor, under the plane of the bed. Cheng et al. [33] also used pressure sensors placed under the legs of a chair to estimate the activity of a person sitting in the chair. They detected seven postural states, including leaning back and crossing one leg over the other knee, with an accuracy of 0.826. However, the force sensors used in these studies are suited for detecting slow changes such as heartbeats, weight shifts, and force-applied movements but have challenges in detecting light, quick movements with weak force, such as swiping on a touch screen. On the other hand, the development of new force sensors with unconventional methods and systems that utilize these sensors is also being actively pursued. Liu et al. [34] have developed a unique soft tube-shaped sensor based on the pneumatic conversion principle as a touch sensor for smart furniture. This sensor is superior in that it can be used on soft furniture materials such as beds, but this method also requires physical deformation of the tube, making it difficult to embed or affix the sensor to nearly non-deformable furniture such as tables. Choi et al. [35] have also developed a force sensor that uses a unique structure to change capacitance linearly in response to pressing pressure. This sensor is excellent for 3D touch sensors because of its high linearity of sensor output in response to applied force. However, because the sensor must be installed on the surface of the panel to be touched, there are issues in using the sensor by hiding it on existing furniture to take advantage of the texture of the material.

E. VIBRATION SENSOR

Kawakatsu et al. [36] proposed an interface embedded in kitchens, washstands and bathrooms that can be operated in wet domestic environments such as kitchens and bathrooms. As an example, they have sensors built into their bathtubs. The embedded sensor is placed behind the bathtub so it cannot be seen from the front. It contains a vibration sensor using a piezoelectric device and an electrostatic sensor to recognize operations such as touching or scratching the bathtub. They distinguish between tapping and scratching by focusing on the fundamental frequency, F0, of the spectrum in the vibration signal specific to the scratching sound. However, they do not consider swiping motions across the bathtub's surface or detecting the direction of such movements.

Yasha et al. [37] are studying an interface using surface acoustic waves (SAWs) using a Voice Pickup Unit (VPU). Since surface acoustic waves propagate on the surface of a material along the boundary between the air and the surface, they are characterized by being less susceptible to ambient noise and propagating over long distances. They have utilized the characteristics of surface acoustic waves to receive the SAWs generated when a VPU sensor is installed on the operating plane of a table and a finger or other object comes into contact with the surface of the table when a gesture is performed, extract features using an optimized Mel Frequency Cepstrum Coefficients (MFCC) for the obtained signal, and classify the type of gesture using machine learning. They have discriminated against seven types of gestures, including no operation, with an accuracy of 95.2%. Also, in an experiment to recognize vibrations associated with cooking actions, such as the operation of cooking utensils on the table, 17 types of actions, including actions that do not operate anything, were classified with a high accuracy of 99.2%. In addition, using two VPU sensors, swipe operations in one axis and two directions are achieved with an accuracy of 99.6%. However, SAWs are constrained by the boundary line between the surface and air and are known to decay exponentially in the depth direction, making it difficult for vibrations to enter the structure. Therefore, when sensors are installed in hidden locations to blend the interface into the room, it is not easy to install them in complex locations not visible from the operating surface, such as behind decorative panels on furniture.

F. HIDDEN INTERFACE

Olwal et al. [25] are researching a hidden display device that realizes on-demand interaction and digital display while using the beauty of familiar materials in furniture and home appliances. This device secures brightness by performing parallel rendering that simultaneously activates and displays the necessary, multiple rows instead of the simple scanning method that sequentially emits light from the light-emitting elements used in general displays. By performing such processing, it is possible to secure 5 to 10 times the brightness of conventional materials, and by embedding this device inside furniture and home appliances and emitting light, characters and symbols can be highlighted on the materials with high brightness. On the other hand, by turning off the light when not in use, only the material can be seen, making it possible to achieve both the beauty and functionality of the material, which was difficult in the past. By embedding these devices in various materials and verifying them, we have realized a display interface that preserves the texture and beauty of materials essential for home interior design. They also conducted a large-scale questionnaire survey considering the realization of a hidden interface and proposed it as a display method that matches the adaptation to aesthetics and user styles. This research is excellent in that it provides new insight into the importance of design in ambient computing, but the focus of the paper is on display devices, and the input interface used in the experiment is a conventional capacitive touch sensor. Capacitive touch sensors have the excellent feature of enabling multi-touch, but in principle, sensitivity decreases in proportion to the distance from the sensor electrodes to the panel surface, making it difficult to use thick materials [38]. In addition, it is generally difficult to use metal materials.

G. SETTING RESEARCH OBJECTIVES VIA EXAMINATION OF RELEVANT LITERATURE

The various sensing and interface research efforts described above are diverse, but they are insufficient to solve issues 1-4 described in Section I and to realize an input interface that blends into the design and aesthetics of the home and the living space. Therefore, in this paper, we try to address these issues.

III. PROPOSED SYSTEM

We propose a stem that enables input operations by touching existing furniture as an input interface that can solve issues 1-4 described in Section I, blends in with the design and aesthetics of homes and living spaces. For the specific input operation, we decided to implement a swipe operation on the furniture's surface. In this section, we first present the requirements that the proposed method must meet and then describe the overall structure of the proposed system. Next, the basic study and detection algorithm for swipe detection will be described in detail.

A. REQUIREMENTS

We have identified the following four requirements for the four issues listed in Section I, and considered their embodiment:

- Requirement 1: Sensing must be possible with a sensor placed in a hidden position, out of sight of the user. (Corresponds to Issue 1)
- Requirement 2: Invasion of privacy must be minimal. (Corresponds to Issue 2)
- Requirement 3: Must be able to operate intuitively, such as being able to move or indicate direction in 4 or more directions. (Corresponds to Issue 3)
- Requirement 4: It is necessary to be able to use textured materials such as metal materials and thick wood for the parts that people see and touch. (Corresponds to Issue 4)

As a sensor to realize a system that meets these requirements, we focused on vibration sensors using piezoelectric elements. Unlike optical cameras and infrared sensors, vibration sensors that use piezoelectric devices are small and thin and can be placed in hidden and invisible positions in furniture and equipment, so they fit Requirement 1. Furthermore, since the sensor captures the vibrations of the user touching the interface instead of images and sounds, there are few privacy concerns, so it matches Requirement 2. Additionally, using deep learning to detect various motions and vibrations that come into contact with furniture and equipment allows for intuitive operation, which is consistent with Requirement 3. Finally, vibrations have the property of being transmitted even through thick materials and metals. For this reason, various materials that emphasize texture can be used for furniture and equipment, and it is possible to increase the degree of freedom in design, which is consistent with Requirement 4.

Table 1 compares existing user interface systems. This table shows that although existing systems can partially realize the above requirements, they do not fulfill all of them. Based on the above discussion, we investigate a method of realizing an input interface that operates existing furniture using a vibration sensor. An operation is to swipe the furniture's surface, and the swipe direction is set to 4 simple directions (UP, DOWN, RIGHT, LEFT). Figure 1 is an image of sensors installed on furniture and indoor walls.

B. OVERVIEW OF THE PROPOSED SYSTEM

Firstly, this section provides an overview of the system. In this study, we chose a table as one of the common pieces of furniture in the home and experimented. Figure 2 shows a block diagram showing the outline of the sensor system of the proposed system, and Figure 3 shows the table used in the experiment and the state of sensor attachment to the table. To detect the direction of swiping the table with a finger in four directions, we thought of capturing the vibration of the table on two axes, the X-axis and the Y-axis, and installed four vibration sensors.

Here is an overview of how the sensor system works, from capturing vibrations on the table to detecting swipes. First, vibrations generated when a person directly touches or swipes a piece of furniture are captured as vibration data using a 4-channel synchronized vibration sensor unit. Each of the four obtained vibration data is converted into data showing temporal changes in frequency components using STFT, and by combining the four data into one array data, data containing correlation information between sensors can be obtained. In addition, the data is labeled according to the swiped direction. Using the data obtained in this way, a learning model is constructed by deep learning, and the constructed model estimates the swipe direction for unknown swipe vibrations. This is followed by a description of the elements that make up the system.

1) BLOCK DIAGRAM OF HARDWARE

Figure 4 shows a block diagram of the hardware section. The vibration signals captured by the four sensors are combined

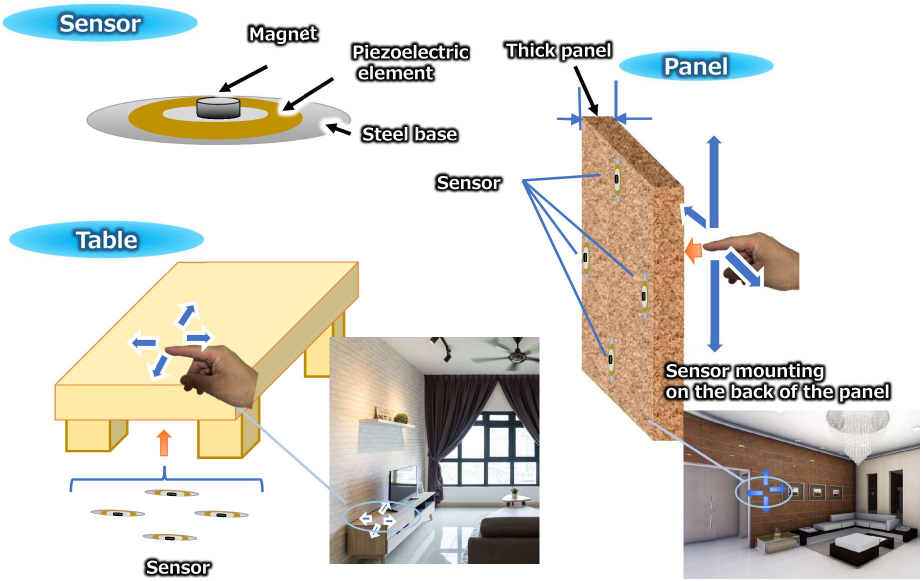

FIGURE 1. Sensor install image.

TABLE 1. Comparison with existing systems related to our work.

	Sensor	Requirement 1	Requirement 2	Requirement 3	Requirement 4
Goto *et al.* [12]	Camera	n/a	n/a	✓	✓
Simone *et al.* [13]	Camera	n/a	n/a	✓	✓
Braun *et al.* [27]	Microphone	✓	n/a	n/a	✓
Mayank *et al.* [28]	Microphone	✓	n/a	n/a	✓
Olwal *et al.* [25]	Electrostatic Touch Sensor	✓	✓	✓	n/a
Pourjafarian *et al.* [29]	Electrostatic Touch Sensor	✓	✓	✓	n/a
Parilusyan *et al.* [30]	Electrostatic Touch Sensor	✓	✓	✓	n/a
Wu *et al.* [31]	Electrostatic Touch Sensor	✓	✓	✓	n/a
Lee *et al.* [32]	Force sensor	✓	✓	n/a	n/a
Cheng *et al.* [33]	Force sensor	✓	✓	n/a	n/a
Liu *et al.* [34]	Force sensor	✓	✓	n/a	n/a
Choi *et al.* [35]	Force sensor	✓	✓	n/a	n/a
Kawakatsu *et al.* [36]	Vibration sensor	✓	✓	n/a	✓
Yasha *et al.* [37]	VPU sensor	n/a	✓	n/a	✓
Our System	Vibration sensor	✓	✓	✓	✓

✓: it fulfills the requirement. n/a: it does not fulfill the requirement.

into two channels for the X-axis (CH1, CH2) and two channels for the Y-axis (CH3, CH4). All channel signals are time-synchronized and recorded in 16-bit Pulse Code Modulation (PCM) format with a sample rate of 44.1 kHz. These synchronized data provide information on how the signals of each sensor channel change relative to each other when a person swipes the tabletop.

2) VIBRATION SENSOR UNIT

We have developed a vibration sensor unit for attaching to furniture. Figure 5 shows the structure of the vibration sensor unit. The sensor consists of a magnetic material such as iron as a base, a piezoelectric device (7BB-41-2L0,[1] Murata Manufacturing, Kyoto, Japan), and a magnet placed in the sensor's center. Devices can be thin, small, low cost, and mounted invisibly behind furniture or even thick decorative panels to capture vibrations from a hidden position. The sensitivity of the fabricated vibration sensor unit depends on the mass of the magnet. We adjusted the mass of the magnet so that it would be sensitive enough to capture a weak signal from a hand swipe and still be well within the dynamic range of the amplifier and the recorder's AD converter.

The magnetic sheet base of the sensor unit can be attached to the back side of furniture or structural materials such as wood using double-sided tape or adhesive. In this study, to detect swipes in four directions, the system detects the direction by synchronously recording the vibration signals from the four sensors and using the relative changes between channels as features.

[1] https://www.murata.com/en-us/api/pdfdownloadapi?cate=cgsubDiaphragms&partno=7BB-41-2L0

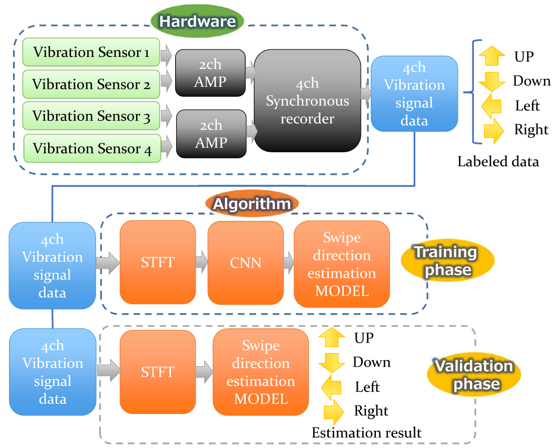

FIGURE 2. System block diagram.

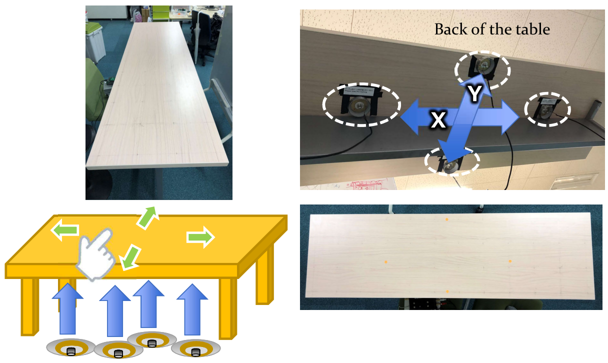

FIGURE 3. Mounting structure of the sensor on a table.

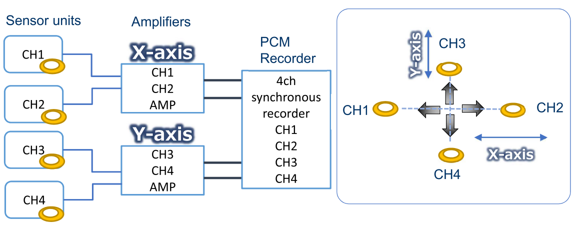

FIGURE 4. Hardware block diagram.

3) SENSOR AMPLIFIER

In a system that captures vibration from the back side of a material and uses that vibration to determine operation, vibration attenuation is an issue, especially in thick materials. In general, it is known that the mechanical vibration of an object attenuates in proportion to the distance from the vibration source. When the distance from the vibration source to the vibration measurement point is r, the amplitude of general mechanical vibration attenuates in the relation $1/r$, and furthermore, near the surface, it attenuates in the relation $1/r^2$. Therefore, the signal level obtained is expected to be small, and low noise is required for the sensor system.

In addition, the dynamic range of motion data, when a person touches a structure, is wide, ranging from large-amplitude signals such as the moment of hand contact to weak signals such as a swipe operation that strokes the surface, so a system with a wide dynamic range that can capture these highly varied signals without failure is required. Therefore, we designed the amplifier circuit focusing on the following two points:

- Ensure a high Signal to Noise Ratio (SNR) of the amplifier by reducing the noise of the circuit
- Ensuring a dynamic range that can capture from minute to large signals without failure

Figure 6 shows a schematic of the main parts of the amplifier circuit. Since the output of a piezoelectric device is generally high impedance, we examined several candidates for Field Effect Transistor (FET) input operational amplifiers with low input bias current and noise characteristics. Furthermore, to achieve a high dynamic range, we designed a dual power supply circuit. Therefore, we chose a rail-to-rail op amp (LTC6241,[2] Analog Devices Inc., Norwood, MA, USA) that satisfies the above requirements, supports dual power supplies, and can swing to the upper limit of the power supply voltage.

In addition, Equation 1 expresses a noise figure of a multi-stage amplifier circuit shown in Figure 7, and expresses the noise figure F_{all} of the entire amplifier circuit when the noise figure of the first stage is F_1, the second stage is F_2, and the nth stage is F_n. As can be seen from the formula, especially to reduce noise in the circuit design, the noise in the first stage of the input section is dominant, so in the first

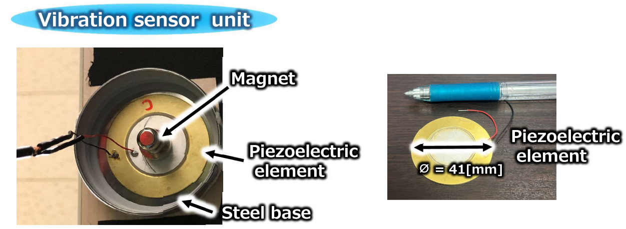

FIGURE 5. Vibration sensor unit.

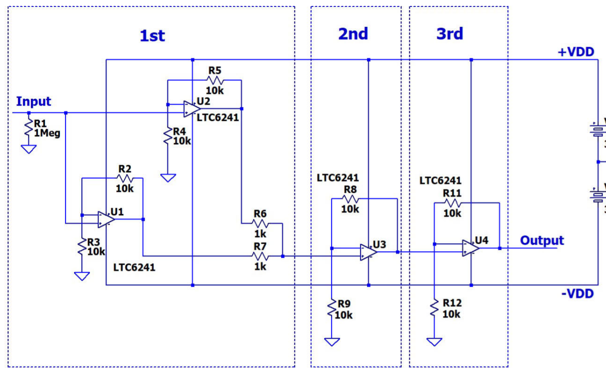

FIGURE 6. Amplifier circuit.

[2]https://www.analog.com/media/en/technical-documentation/data-sheets/624012fe.pdf

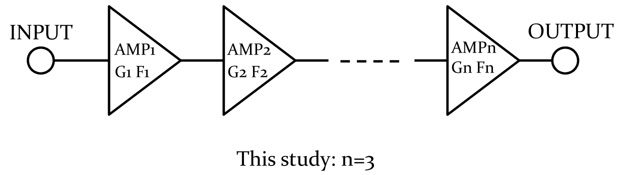

FIGURE 7. Multi-stage amplifier circuit.

stage amplifier circuit of the input section, which is connected immediately after the output of the vibration sensor unit, two operational amplifiers are used in parallel to reduce the noise generated in the amplifier circuit section to $1/\sqrt{2}$.

$$F_{all} = F_1 + \frac{F_2}{G_1} + \frac{F_3}{G_1 \cdot G_2} + \cdots + \frac{F_n - 1}{G_1 \cdot G_2 \cdots G_n} \tag{1}$$

The amplifier produced in this research has a three-stage configuration. Since the quantization bit number of the PCM recorder is 16 bits, the target SNR for the amplifier circuit was set to 105dB or more, which is slightly higher than the ideal SNR of a 16-bit AD converter, which is approximately 98.08 dB. The configuration of the amplifier is such that the first-stage amplification circuit is used for impedance conversion, and the amplification degree can be adjusted in the second and third stages. The reasons for the second and third-stage amplification circuits are the following. At the time of the initial study, it was unclear how much amplification would be required to detect the vibration caused by swiping. For this reason, multiple amplification circuits were installed in series to accommodate a wide range of amplification levels and to maintain as much negative feedback per amplifier circuit as possible to ensure SNR. The amplifier used in this study was designed with a total gain of approximately 18 dB and a maximum output amplitude of 6 dBV, and the measured SNR was 112.5 dB (with a 30 kHz Low-Pass Filter).

By implementing these measures, it is possible to reliably amplify signals from subtle vibrations during swipes to significant vibrations when touched, all with low noise, in various devices.

4) ENCLOSURE DESIGN FOR LOW NOISE

In indoor sensing, hum noise from power lines and unnecessary radiation from other home appliances become a problem in practical use, and these noises reduce the S/N ratio of amplifiers. This system has a particularly low sensor signal level and uses a circuit with high impedance to handle piezoelectric devices, so in addition to reducing the noise of the circuit itself, it is necessary to take countermeasures against external noise. Therefore, to protect the cables of the system and the amplifier board from radiated and conducted noise in the home, an aluminum shield was used for the housing of the sensor amplifier system, and the cables were of coaxial construction, as shown in Figure 8. In addition, by mounting the amps of the channels in which the difference between the two channels is expected to be important in one housing, the conditions for the amp circuits of the two channels are aligned.

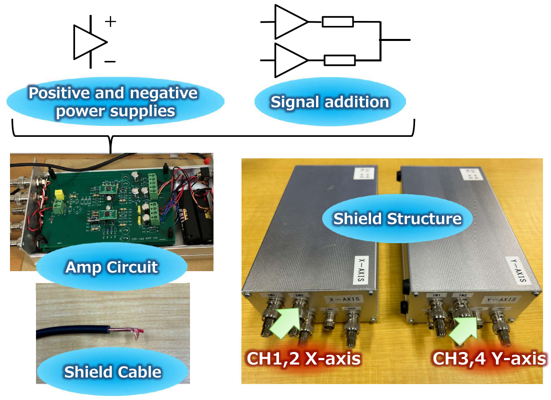

FIGURE 8. Amplifier system.

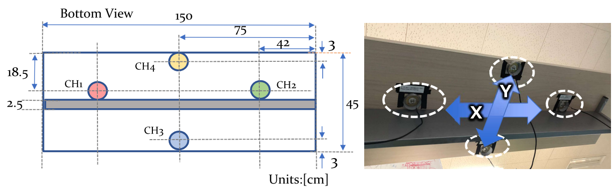

FIGURE 9. Mounting position of the sensor.

C. FEATURES OF SIGNAL WHEN SWIPING

We used the aforementioned system to analyze the data to see how the sensors could receive vibrations from swiping on the table. These experiments were conducted to identify the features used to detect the direction of the swipe on the table. This section describes what we found. Figure 9 shows the mounting structure of the sensor on a table, with four vibration sensors attached to the bottom of the 150 cm table. The swipe directions are swipe from left to right (Right), swipe from right to left (Left), swipe from bottom to top (Up), and swipe from top to bottom (Down). To detect a swipe in four directions, two vibration sensor units are used for each of the horizontal X-axis and vertical Y-axis from the swiping person's point of view, and a total of four vibration sensor units are used for each element. The goal was to use vibration data to determine the direction in which a person swiped.

1) OBSERVATION OF VIBRATION DATA TIME AXIS WAVEFORMS

Figure 10 shows the signal waveform data when a person swiped the top board with a finger in the above system. Initially, we thought it would be possible to easily detect finger movement from the relative amplitude level difference between the start and end points of a swipe. For example, we guessed that moving a finger from near the ch1 sensor to near the ch2 sensor would initially result in the amplitude of ch1 being larger than that of ch2, and the amplitude of ch2

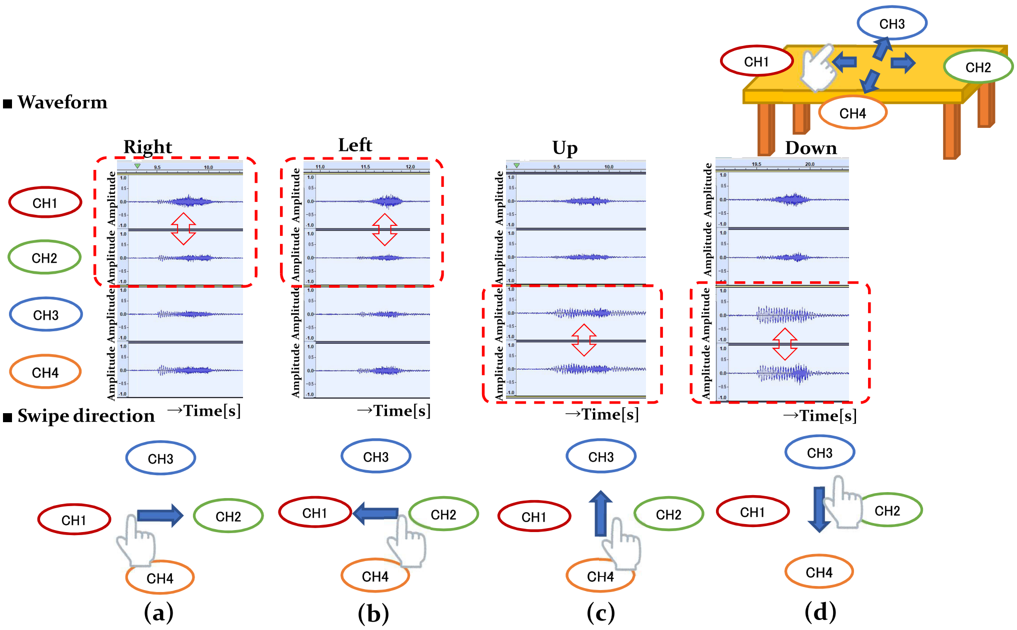

FIGURE 10. Waveform data.

would become larger than that of ch1 gradually. However, when confirming the actual signal, it was impossible to confirm the bias in the clear size relationship of the amplitude for each axis channel going with the movement of the finger swiping.

2) VIBRATION SOURCE LOCATION AND SENSOR SIGNAL AMPLITUDE

Based on the above results, we investigated in detail the position where vibration was applied to the table and the output amplitude level of each sensor. To investigate the relationship between the distance from the vibration source and the output of the vibration sensor in detail, we experimented using a vibrator that generates vibration, as shown in Figure 11. As shown in the diagram, straight lines were drawn dividing the tabletop surface at equal intervals of 130 mm in length and width, and 48 points of intersection were excited one by one with a vibrator, and the sensor output level at that time was recorded. The excitation frequency is a sinusoidal signal of six frequencies ranging from 100 Hz to 4 kHz.

Figure 12 shows the relationship between the distance between the sensor and the vibrator and the sensor output for 400 Hz among the results. The graphs are shown separately for each channel, CH1 to CH4. As seen from Figure 12, there is no simple and clear trend that the sensor output level decreases in proportion to the distance from the vibrator to the sensor, but rather the amplitude increases in some cases as the distance from the vibrator.

Next, we checked the characteristics at each signal frequency. Figure 13 shows the relationship between the distance between the CH4 sensor and the vibrator out of the four sensors and the sensor output for each frequency. No clear trend of decreasing sensor signal level inversely proportional to distance could be found at any frequency. Experiments also showed that the sensor output level may increase rapidly when a specific position is vibrated at a specific frequency due to table resonance and other effects. Thus, it was found that it is difficult to identify the position and movement of the vibration source using only simple amplitude levels.

3) AUDITORY CONFIRMATION

In the case of sound waves, it is known that the position and movement of a sound source can be detected by using information on changes in frequency components and phase differences between channels in addition to amplitude information. To ascertain the applicability of a similar approach to vibration, we used the fact that the frequency band of the signal of each sensor obtained from the vibration sensor is the voice band. Specifically, we verified the results by playing back the PCM data recorded by swiping in four directions, up, down, left, and right, with a recorder and then playing it back as audible sounds with headphones.

As shown in Figure 14, the X-axis CH1 was applied to the left ear and CH2 to the right ear, and the vibration data was aurally confirmed. Similarly, for the Y axis, CH3 was added to the left ear, and CH4 was added to the right ear, and the vibration was confirmed as sound aurally. As a result, we could perceive changes in the frequency components contained in the audible sound as the finger's movement during the swipe operation. Just to be sure, similar confirmation was performed on several other tables, and the same results were obtained regarding auditory perception.

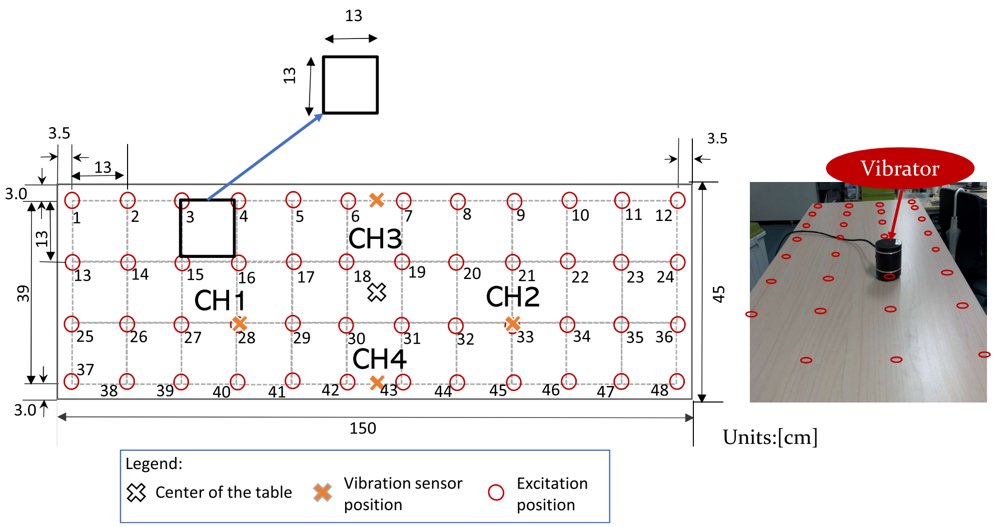

FIGURE 11. Measurement of the sensor signal amplitude at each position of the vibration source.

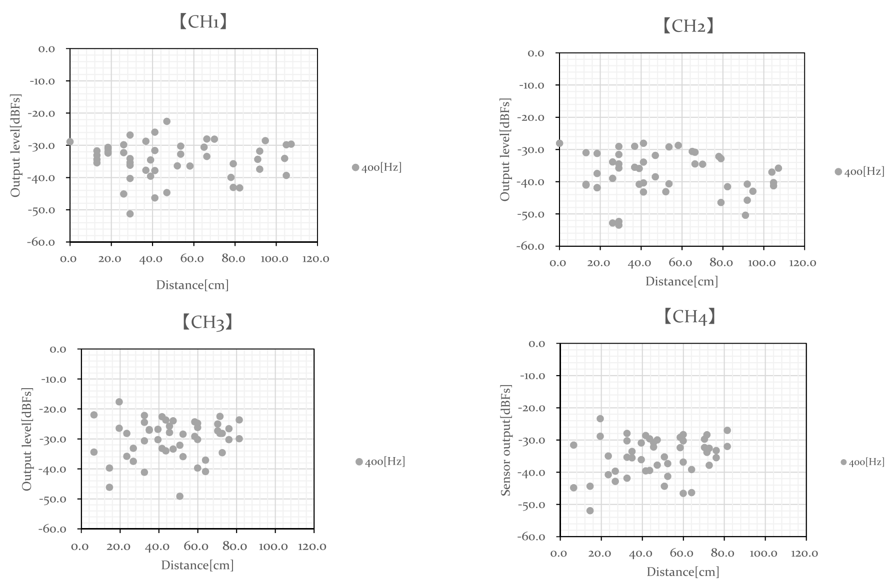

FIGURE 12. Distance between vibration sensor and vibration source VS sensor output(400Hz).

4) CONFIRMATION USING SPECTROGRAM

Since the direction of finger movement during swiping was audibly recognizable, the signal obtained was confirmed as a spectrogram, as shown in Figure 15 to clarify the reason. (a) to (d) in Figure 15 are spectrograms under the following conditions:

(a) Swipe from left to right.
(b) Swipe from right to left.
(c) Swipe from bottom to top.
(d) Swipe from top to bottom.

The vertical axis is frequency, the horizontal axis is time, and the FFT (Fast Fourier Transform) is performed for each channel so that changes in the distribution of the frequency components of each channel over time can be confirmed.

In Figure 15(a), for instance, the swipe is made from left to right, the finger initially starting close to the CH1 sensor and ending close to the CH2 sensor. In these circumstances, as indicated by the arrows in the figure, the spectrograms of CH1, which correspond to the start position, and CH2, which correspond to the stop position, can be confirmed as an upward-sloping striped pattern and a downward-sloping pattern, respectively. When the spectrograms of Figures 15(b) to (d) were confirmed in the same manner, the same tendency was observed, although it was difficult to distinguish the striped pattern in some parts. That is, the spectrogram of the channel's signal near the position where the swipe is started has an upward slope and the signal near the position where the swipe stops have a downward slope. It was found that these signal changes are the same as

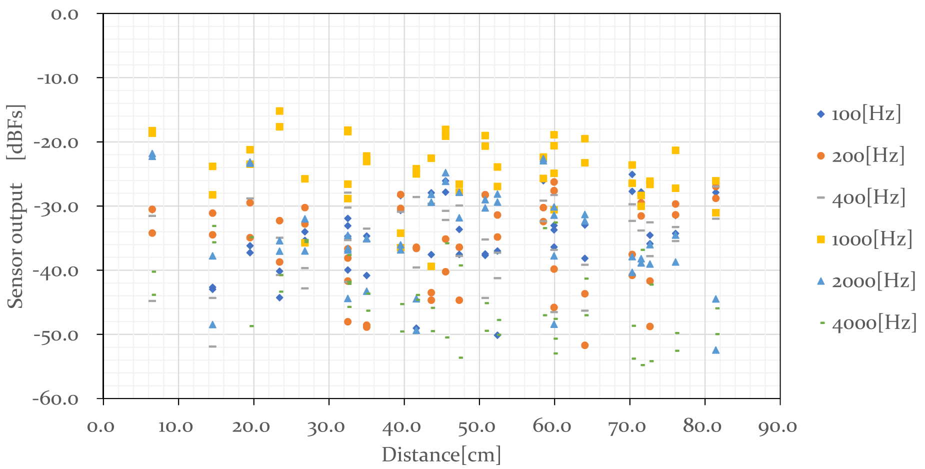

FIGURE 13. Distance between vibration sensor and vibration source VS sensor output(6 different frequencies).

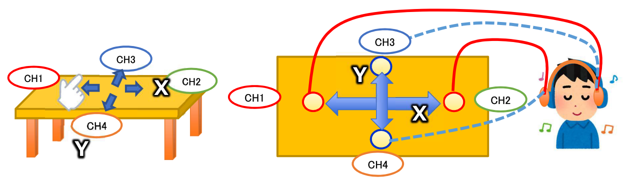

FIGURE 14. Sound confirmation.

TABLE 2. STFT parameters before and after improvement.

Parameter	Previous study [26]	This study
nperseg	1000	4096
noverlap	500	2048
nfft	1000	4096
Timeresolution[mSec]	11.34	46.44
Frequency resolution[Hz]	44.1	10.8

changes in "harmonics" composed of the fundamental wave and harmonics when viewed as temporal changes in acoustic signals, which is why the human ear can perceive them.

However, the auditory confirmation here is based on the waveform of the vibration when the hand touches the table, which is converted into sound. Because it is not a direct recording of sound waves transmitted through the air like a microphone, this method is characterized by less concern for privacy violations.

Based on these results, we studied using deep learning to learn the vibration signal as a temporal change of the spectrogram and determine the swipe direction.

D. SWIPE DIRECTION DETECTION USING DEEP LEARNING

In this section, we will explain swipe direction detection using deep learning together with the preprocessing process.

1) PREPROCESSING

The preprocessing for feature extraction is performed as follows. Pre-process the data obtained by swiping in the four directions (Right, Left, Up, and Down) as follows:

(1) Cut out 2 seconds of data, including silence before and after the swipe, as data for one swipe.
(2) Transform the data of each of the channels CH1-CH4 by STFT (Short-Time Fourier Transform) (Han window is used for the window function) and create an array.
(3) Concatenate four channel arrays into one.
(4) labeling the concatenated data in four directions (Right, Left, Up, Down) to indicate which direction the data was swiped from.

We had similarly performed the above pre-processing in our previous study [26] but had set the number of STFT segments at 1000 in Step 2. However, reconsidering the evaluation results of the auditory perception experiment presented in the previous section, it is thought that when people hear the swipe vibration, they discriminate the swipe direction as a result of hearing the fine pitch changes in the fundamental and harmonic components. This suggests that the frequency component is the most important information in determining direction, and in this paper, the pre-processing of the vibration data during swiping was reviewed again, and the parameters were reconsidered to capture the changes in the frequency axis in more detail. As a result, the number of segments was changed from 1000 to 4096, as shown in Table 2. This change improved the frequency resolution by a factor of approximately four.

2) SWIPE DIRECTION LEARNING AND DETECTION USING CNN

In our proposed method, vibration data from four channels of sensors are each converted into a spectrogram by STFT, and then the results of the four channels are combined into a single array to produce data that looks like a single image. For this reason, we considered using a CNN, commonly used

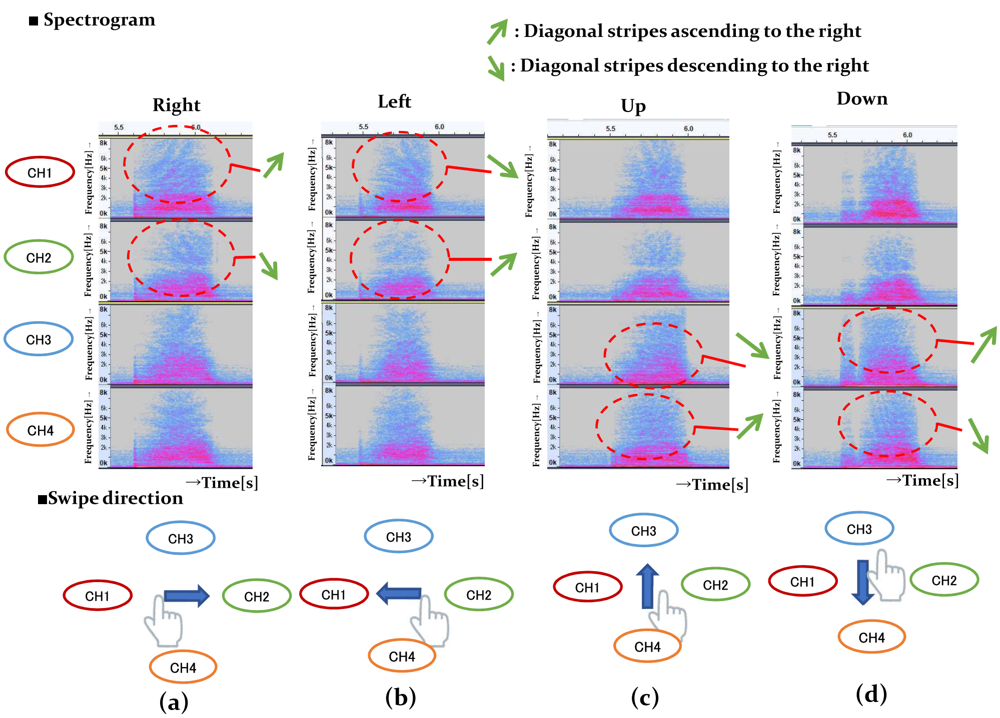

FIGURE 15. Spectrogram.

in imaging, for classification. In this study, we implemented the deep learning model with Keras,[3] a deep learning library for Python. In our previous study [26], we examined learning and classification with a relatively simple deep learning model such as Figure 16(a). However, when the data to be discriminated was the same as the learned person, the discrimination was highly accurate, exceeding 0.9, but the classification accuracy was insufficient in the LOPOCV. Therefore, we tried configurations such as increasing the number of layers in the CNN and adding new preprocessing to improve the accuracy further. As a result, as shown in Figure 16 (b), the number of layers of the CNN used in this study was set to seven layers, including the convolutional and all-junction layers, and a residual connection structure was also introduced.

3) DEEP LEARNING MODELS

The input data to the CNN model configured as shown in Figure 16 (b) has a size of (4098, 90), which is a composite of the STFT results for four channels merged into a single 2D array. The data is input through an input layer tuned to the number of arrays corresponding to the size. After that, two convolutional layers of size (3×3) with 64 filters, each using ReLU as an activation function, are connected, followed by a max-pooling layer (2×2). Then, two convolution layers of convolution size (3 × 3) with 64 filters are used as a Residual Connection. Subsequently, calculations are performed using

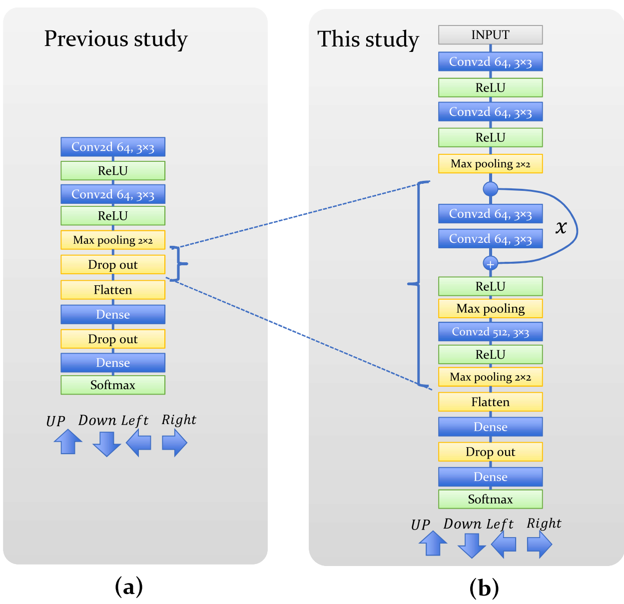

FIGURE 16. Layer structure of CNN.

the activation function ReLU, followed by the max-pooling layer (2×2), and convolution is performed using 512 filters of convolution size (3×3). Then, perform calculations using the activation function ReLU, followed by a max-pooling layer (2×2). Then, after the Flatten process is performed, fully connected layers, a Dropout layer, and the next set of fully connected layers are passed, and finally, the softmax

[3] https://keras.io/

function is used to output the decision results for the four directions (UP, DOWN, LEFT, and RIGHT) corresponding to the direction in which the person swiped. The major changes compared to previous methods are as follows:

(1) The convolution layers increased from two to five.
(2) A Residual Connection was added to the third and fourth convolutional layers.
(3) The number of convolution filters in the fifth layer increased to 512.

He et al. [39] introduced Residual Connections, an architecture commonly utilized in networks with 20 or more layers. This architecture was employed in our model because its effectiveness was verified in initial experiments, in which the number of layers was gradually increased by one layer.

The learning model using the CNN was created with a batch size of 10 and 48 learning epochs, using the loss function 'categorical cross entropy' and the optimization algorithm 'adam.'

IV. SYSTEM EVALUATION AND RESULTS

This section describes the evaluation results. We used the proposed method to evaluate the detection accuracy when swiping a finger in four directions (up, down, left, and right) on a table using data from three participants. Learning the swipe direction using CNN and evaluating the accuracy was performed on a PC with the following specifications:

- CPU: 12th Gen Core i7 (12700H) 2.3GHz
- Memory: 16GB DDR5 4800MHz
- Storage: 1TB
- OS: Windows 11

The evaluation was conducted in two ways: by increasing the number of STFT segments from 1000 to 4096 (New method 1) and improving the CNN structure after increasing the number of STFT segments from 1000 to 4096 (New method 2).

A. DATASET

Three participants (Participant A, Participant B, and Participant C) participated in the study, as shown in Figure 17. Data for each session were acquired by swiping 10 times in each of the four directions (up, down, left, and right), resulting in the collection of nine sessions of data per person. As a result, a total of 27 sessions of data were obtained. The experimental environment was conducted in an office/laboratory environment with people walking nearby and conversing.

B. LEAVE ONE PERSON OUT CROSS VALIDATION

First, as shown in Figure 18, the evaluation uses LOPOCV, the data of 2 out of 3 participants as learning data, and the data of the remaining 1 participant as evaluation data. In order to create a learning model for evaluating this method under the same conditions as the previous study [26], three sessions of data per person were used for the evaluation. This evaluation allows us to see how well a model trained on data other than the person using the interface can detect the swiping actions of a person using the interface.

The results are shown in the "Leave-One-Person-Out-Cross-Validation" row of Table 3. As shown in the table, when the frequency resolution was increased by changing the STFT parameters, the accuracy improved by approximately 4% compared to the previous study [26], resulting in an accuracy value of 0.56. Furthermore, by devising the deep learning layer structure, the accuracy improved by approximately 15%. As a result, the accuracy value improved to 0.67.

C. CONFIRMING THE ACCURACY WHEN ADDING ONE SESSION OF THE TARGET PERSON'S DATA TO THE DATASET USED IN LOPOCV

Furthermore, experiments were conducted assuming the case where the user tunes the interface with their data in advance. We confirmed the accuracy of a trained model by creating it with the addition of only one session of evaluation target data to each learning dataset used in the LOPOCV, as shown in Figure 19. The purpose of this experiment was to see if creating a model with a little additional data on the users who use the interface would improve accuracy, an experiment designed for real-world operation.

The results are shown in the "Adding only one session of target person data" row of Table 3. Looking at the results, when we added only one session's worth of swipe data of the person whose swipe direction was to be estimated when compared to the previous study [26], changing the STFT parameters improved the accuracy value by approximately 1% to 0.75. Furthermore, by improving the layer structure of deep learning, the accuracy value improved by approximately 16%, resulting in an accuracy value of 0.90.

From these results, we obtained sufficient accuracy for practical use as an interface, especially if the user tunes in advance. As a result, it is possible to realize an interface that allows users to touch and operate furniture made of materials such as thick wood directly, which is difficult to achieve with an electrostatic touch panel.

V. DISCUSSIONS

A. IMPLEMENTATION OF VIBRATION SENSORS ON FURNITURE

Vibrations are transmitted through materials. Therefore, even when vibration sensors are installed in inconspicuous locations such as behind interior decorative panels, furniture tops, and walls, they can capture various information in vibrations caused by human contact and impact. By taking advantage of this property, a system that uses vibration sensors hidden in invisible areas of furniture and other interior components can be constructed to make the most of the materials without compromising the interior design, thereby realizing an interface that blends in with daily life. In addition, vibration sensors do not directly record sound waves transmitted through the air like microphones do,

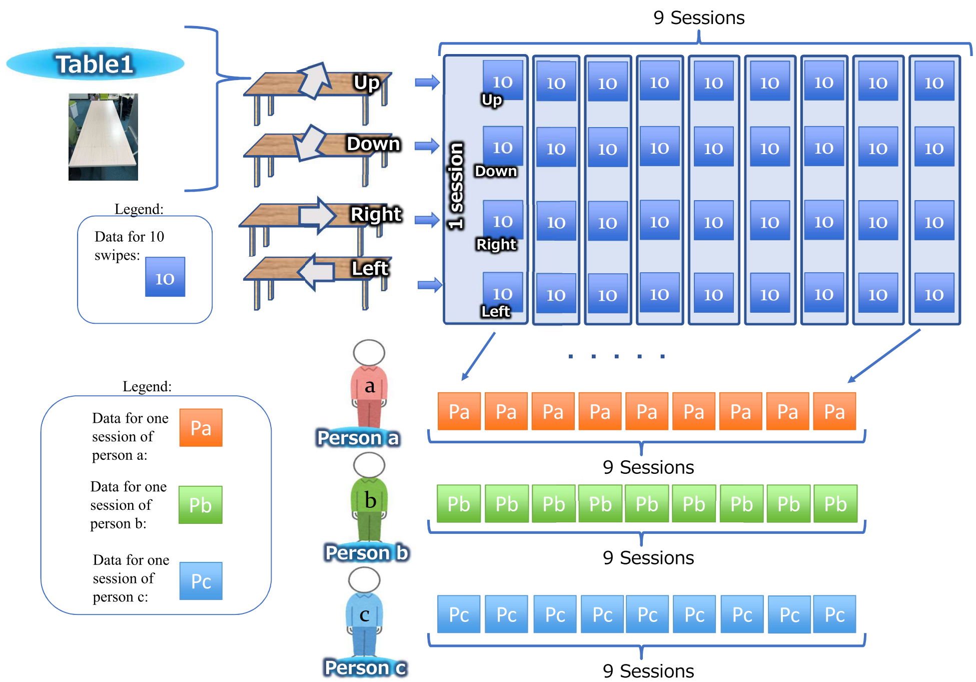

FIGURE 17. Dataset for each person.

TABLE 3. Accuracy of evaluation result.

Evaluation condition	Previous study [26]	New Method1	New Method2
Leave-One-Person-Out-Cross-Validation	0.52	**0.56**	**0.67**
Confirming the accuracy when adding one session of the target person's data to the dataset used in Leave-One-Person-Out-Cross-Validation	0.74	**0.75**	**0.90**

New Method1:Increased the number of STFT segments from 1000 to 4096.
New Method2:Increased the number of STFT segments from 1000 to 4096 and adopted a new CNN structure.

a feature that reduces concerns about privacy violations. Furthermore, vibration sensors have the advantage of being able to sense from the back side of thick materials and metals, which is difficult with conventional electrostatic sensors. They are a powerful sensing method when considering ambient interfaces.

However, unlike sound waves traveling through the air at a constant speed, vibration waves traveling through furniture and building materials made up of various members with different densities are detected as composite waves of vibrations traveling through multiple paths with different propagation speeds. Therefore, it is very difficult to estimate the position and movement of the vibration source by TDoA (Time Difference of Arrival) using the arrival time of the vibration or by AoA (Angle of Arrival) using the angle of arrival.

We overcame this challenge by synchronizing multiple channels of sensors to capture changes in the frequency components of vibrations at the start and stop of a swiping operation by touching an object and using deep learning to learn and identify the operation. Also, to realize an interface using vibration, there was a problem in that it had to be able to capture signals without failure, from large vibrations, such as the moment a hand touches, to weak vibrations, such as swiping. In this research, we have developed a dedicated low-noise amplifier using a positive and negative dual power supply system to obtain a large dynamic range in response to this problem. Furthermore, by devising a layered structure of preprocessing and deep learning, we have improved the accuracy to 0.90 and built a practical system, although user tuning is required.

In addition, in the report of this experiment, only a wooden table was used, but as an additional test, when the vibration of swiping on the metal material was confirmed as a sound, a similar change in tone was confirmed. Therefore, even if the furniture is made of metal, it will be possible to detect the swipe direction using vibration data. Such a system for detecting changes in the movement of objects using vibration contributes to the realization of design-oriented facilities and equipment that can be operated by touch, taking advantage of the aesthetics and texture of the materials.

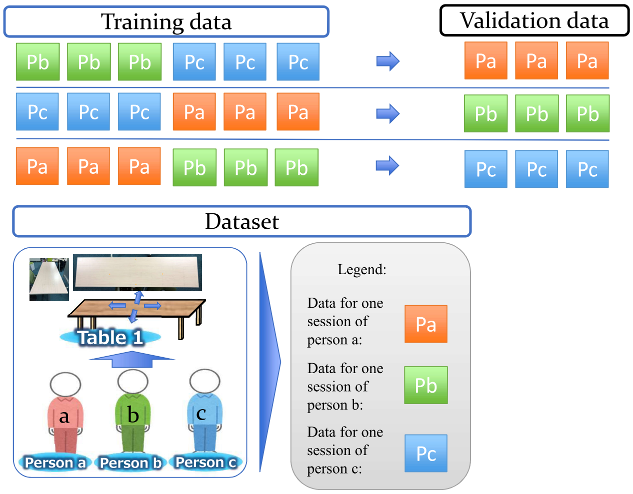

FIGURE 18. Leave-one-person-out-cross-validation.

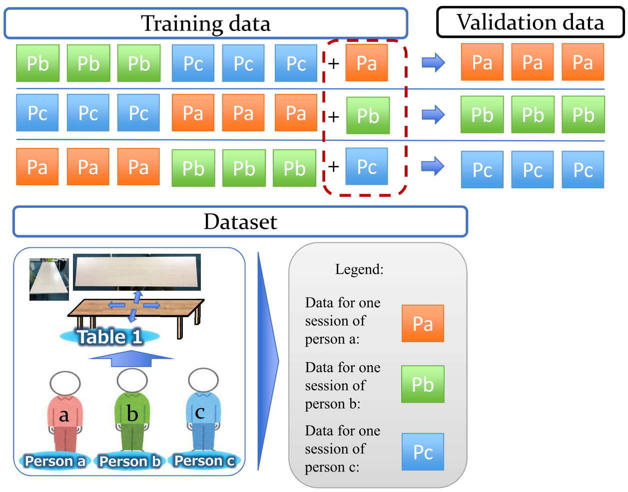

FIGURE 19. Added to one session to training data.

However, detecting vibrations in furniture, such as cushions and sofas that absorb vibrations, is difficult due to the mechanism of the vibration sensor. To solve this problem, we think it is necessary to consider a hybrid structure that combines soft materials, such as providing a non-deformable solid base member that is integrated with the piezoelectric device.

B. SYSTEM PROCESSING SPEED AND IN-HOME IMPLEMENTATION

In this study, vibration data was trained using CNN on a PC with the specifications described in Section IV, using the CPU without a GPU. As an example of the time required to build a training model, the time required to build a training model by training data for six sessions was approximately 7,611 seconds. We also measured the time directly related to the response when using the interface. As a result, it took about 0.015 seconds to extract features using STFT and about 0.35 seconds to determine the swipe. 0.365 seconds, which is the added value of these times, is slower than the response time of a typical touch panel but is adequate for operating furniture by touching it. However, if a CPU with lower performance, such as that used in edge computing, is used, the response is expected to be a little slower, and further elaboration is needed.

Next, we examined the actual use of this system in the home. There are two possible ways to implement this research as a system for actual use with furniture in the home. The first is to detect the collected vibration data with a small edge computing device placed near the furniture, such as a table. We consider it possible to build the training model on an edge computer due to the small scale of the CNN model used in this study, but if the time required for training is unacceptable, training could be performed in the cloud via Wi-Fi. In this case, too, vibration data has the advantage that, unlike camera video data, the amount of data is small and can be transferred at high speed. The other method is to wirelessly transmit vibration data from the sensor unit to the in-home network and perform learning and swipe direction determination on the cloud side. Although propagation delay time is an issue, this method will become feasible as network communication speeds and delays improve with the development of network systems.

VI. CONCLUSION

In this paper, we investigated improving detection accuracy for an input interface system using vibration data during swiping. The research system consists of four time-synchronized vibration sensor units and a high-dynamic-range, low-noise amplifier to capture the vibrations when a person swipes a table. Then, the temporal change in the frequency of the captured vibration was input as a patterned image-like array using a spectrogram and was recognized using deep learning.

In the evaluation of the system, accuracy was confirmed by LOPOCV using Three sessions of swiped data for each individual for 3 participants, and the accuracy was 0.67, which is approximately 15% higher than the red previous study [26]. Furthermore, in order to validate the accuracy when users pre-tune the interface with their own data, we examined the accuracy of the model trained by adding only one session of evaluation target data to each learning dataset used in the aforementioned LOPOCV. As a result, the accuracy reached 0.90, demonstrating an approximately 16% improvement in accuracy compared to the previous study [26], and practical accuracy was achieved.

In the future, we plan to examine ways to improve accuracy by using more training data, adding operations other than swiping, and examining methods for retrying false positives in order to develop the system as a practical one. In addition, although we focused on tables as furniture this time, we plan to conduct similar studies on other furniture and walls of buildings.

REFERENCES

[1] O. Omojola, E. R. Post, M. D. Hancher, Y. Maguire, R. Pappu, B. Schoner, P. R. Russo, R. Fletcher, and N. Gershenfeld, "An installation of interactive furniture," *IBM Syst. J.*, vol. 39, no. 3.4, pp. 861–879, 2000.

[2] O. Krejcar, P. Maresova, A. Selamat, F. J. Melero, S. Barakovic, J. B. Husic, E. Herrera-Viedma, R. Frischer, and K. Kuca, "Smart furniture as a component of a smart city—Definition based on key technologies specification," *IEEE Access*, vol. 7, pp. 94822–94839, 2019.
[3] Z. Vlaović, M. Jaković, and D. Domljan, "Smart office chairs with sensors for detecting sitting positions and sitting habits: A review," *Drvna industrija*, vol. 73, no. 2, pp. 227–243, May 2022.
[4] R. Frischer, O. Krejcar, P. Maresova, O. Fadeyi, A. Selamat, K. Kuca, S. Tomsone, J. P. Teixeira, J. Madureira, and F. J. Melero, "Commercial ICT smart solutions for the elderly: State of the art and future challenges in the smart furniture sector," *Electronics*, vol. 9, no. 1, p. 149, Jan. 2020. [Online]. Available: https://www.mdpi.com/2079-9292/9/1/149
[5] O. Patsadu, C. Nukoolkit, and B. Watanapa, "Human gesture recognition using Kinect camera," in *Proc. 9th Int. Conf. Comput. Sci. Softw. Eng. (JCSSE)*, May 2012, pp. 28–32.
[6] F. Erden and A. E. Çetin, "Hand gesture based remote control system using infrared sensors and a camera," *IEEE Trans. Consum. Electron.*, vol. 60, no. 4, pp. 675–680, Nov. 2014.
[7] Z. Ren, J. Meng, and J. Yuan, "Depth camera based hand gesture recognition and its applications in human-computer-interaction," in *Proc. 8th Int. Conf. Inf., Commun. Signal Process.*, Dec. 2011, pp. 1–5.
[8] C. Wang, Z. Liu, and S.-C. Chan, "Superpixel-based hand gesture recognition with Kinect depth camera," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 29–39, Jan. 2015.
[9] Z. Ren, J. Yuan, and Z. Zhang, "Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera," in *Proc. 19th ACM Int. Conf. Multimedia (MM)*. New York, NY, USA: Association for Computing Machinery, Nov. 2011, pp. 1093–1096, doi: 10.1145/2072298.2071946.
[10] S. S. Ge, Y. Yang, and T. H. Lee, "Hand gesture recognition and tracking based on distributed locally linear embedding," *Image Vis. Comput.*, vol. 26, no. 12, pp. 1607–1620, Dec. 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885608000693
[11] J. Galván-Ruiz, C. M. Travieso-González, A. Tejera-Fettmilch, A. Pinan-Roescher, L. Esteban-Hernández, and L. Domínguez-Quintana, "Perspective and evolution of gesture recognition for sign language: A review," *Sensors*, vol. 20, no. 12, p. 3571, Jun. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/12/3571
[12] H. Goto, D. Takemura, Y. Kawasaki, and A. Nakamura, "Development of an information projection interface using a projector-camera system," *Electron. Commun. Jpn.*, vol. 96, no. 11, pp. 70–81, 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ecj.11477
[13] S. Pizzagalli, D. Spoladore, S. Arlati, M. Sacco, and L. Greci, "HIC: An interactive and ubiquitous home controller system for the smart home," in *Proc. IEEE 6th Int. Conf. Serious Games Appl. Health (SeGAH)*, May 2018, pp. 1–6.
[14] M. Tsuji, H. Kubo, S. Jayasuriya, T. Funatomi, and Y. Mukaigawa, "Touch sensing for a projected screen using slope disparity gating," *IEEE Access*, vol. 9, pp. 106005–106013, 2021.
[15] F. Portet, M. Vacher, C. Golanski, C. Roux, and B. Meillon, "Design and evaluation of a smart home voice interface for the elderly: Acceptability and objection aspects," *Pers. Ubiquitous Comput.*, vol. 17, no. 1, pp. 127–144, Jan. 2013.
[16] W. Mao, J. He, and L. Qiu, "CAT: High-precision acoustic motion tracking," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*. New York, NY, USA: Association for Computing Machinery, Oct. 2016, pp. 69–81, doi: 10.1145/2973750.2973755.
[17] K. Sun, T. Zhao, W. Wang, and L. Xie, "VSkin: Sensing touch gestures on surfaces of mobile devices using acoustic signals," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*. New York, NY, USA: Association for Computing Machinery, Oct. 2018, pp. 591–605, doi: 10.1145/3241539.3241568.
[18] M. Schrapel, M.-L. Stadler, and M. Rohs, "Pentelligence: Combining pen tip motion and writing sounds for handwritten digit recognition," in *Proc. CHI*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–11, doi: 10.1145/3173574.3173705.
[19] H. Yin, A. Zhou, G. Su, B. Chen, L. Liu, and H. Ma, "Learning to recognize handwriting input with acoustic features," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 2, pp. 1–26, Jun. 2020, doi: 10.1145/3397334.
[20] Q. Wan, Y. Li, C. Li, and R. Pal, "Gesture recognition for smart home applications using portable radar sensors," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 6414–6417.
[21] M. Al-Qaness and F. Li, "WiGeR: WiFi-based gesture recognition system," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 6, p. 92, Jun. 2016. [Online]. Available: https://www.mdpi.com/2220-9964/5/6/92
[22] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using WiFi," in *Proc. MobiSys*. New York, NY, USA: Association for Computing Machinery, 2017, pp. 252–264, doi: 10.1145/3081333.3081340.
[23] C. Li, M. Liu, and Z. Cao, "WiHF: Enable user identified gesture recognition with WiFi," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Jul. 2020, pp. 586–595.
[24] H. Abdelnasser, K. Harras, and M. Youssef, "A ubiquitous WiFi-based fine-grained gesture recognition system," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2474–2487, Nov. 2019.
[25] A. Olwal and A. Dementyev, "Hidden interfaces for ambient computing: Enabling interaction in everyday materials through high-brightness visuals on low-cost matrix displays," in *Proc. CHI Conf. Hum. Factors Comput. Syst.* New York, NY, USA: Association for Computing Machinery, 2022, pp. 1–20, doi: 10.1145/3491102.3517674.
[26] M. Yoshida, T. Matsui, T. Ishiyama, M. Fujimoto, H. Suwa, and K. Yasumoto, "Smatable: A system to transform furniture into interface using vibration sensor," in *Proc. 19th Int. Conf. Intell. Environ. (IE)*, Jun. 2023, pp. 1–8.
[27] A. Braun, S. Krepp, and A. Kuijper, "Acoustic tracking of hand activities on surfaces," in *Proc. 2nd Int. Workshop Sensor-based Activity Recognit. Interact. (iWOAR)*. New York, NY, USA: Association for Computing Machinery, Jun. 2015, pp. 1–5, doi: 10.1145/2790044.2790052.
[28] M. Goel, B. Lee, M. T. I. Aumi, S. Patel, G. Borriello, S. Hibino, and B. Begole, "SurfaceLink: Using inertial and acoustic sensing to enable multi-device interaction on a surface," in *Proc. SIGCHI Conf. Human Factors Comput. Syst. (CHI)*. New York, NY, USA: Association for Computing Machinery, Apr. 2014, pp. 1387–1396, doi: 10.1145/2556288.2557120.
[29] N. Pourjafarian, A. Withana, J. A. Paradiso, and J. Steimle, "Multi-touch kit: A do-it-yourself technique for capacitive multi-touch sensing using a commodity microcontroller," in *Proc. 32nd Annu. ACM Symp. User Interface Softw. Technol. (UIST)*. New York, NY, USA: Association for Computing Machinery, Oct. 2019, pp. 1071–1083, doi: 10.1145/3332165.3347895.
[30] B. Parilusyan, M. Teyssier, V. Martinez-Missir, C. Duhart, and M. Serrano, "Sensurfaces," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 2, pp. 1–19, Jul. 2022, doi: 10.1145/3534616.
[31] T. Wu, S. Fukuhara, N. Gillian, K. Sundara-Rajan, and I. Poupyrev, "ZebraSense: A double-sided textile touch sensor for smart clothing," in *Proc. 33rd Annu. ACM Symp. User Interface Softw. Technol. (UIST)*. New York, NY, USA: Association for Computing Machinery, Oct. 2020, pp. 662–674, doi: 10.1145/3379337.3415886.
[32] W. Lee, H. Yoon, C. Han, K. Joo, and K. Park, "Physiological signal monitoring bed for infants based on load-cell sensors," *Sensors*, vol. 16, no. 3, p. 409, Mar. 2016. [Online]. Available: https://www.mdpi.com/1424-8220/16/3/409
[33] J. Cheng, B. Zhou, M. Sundholm, and P. Lukowicz, "Smart chair: What can simple pressure sensors under the chairs legs tell us about user activity," in *Proc. 7th Int. Conf. Mobile Ubiquitous Comput., Syst., Services Technol. (UBICOMM)*, 2013, pp. 81–84.
[34] S. L. Liu, H. Cai, and C. Liu, "Soft body belt-type touch sensor with impact resistance: A study of dynamic behavior," *IEEE Access*, vol. 9, pp. 128460–128466, 2021.
[35] H. B. Choi, J. Oh, Y. Kim, M. Pyatykh, J. C. Yang, S. Ryu, and S. Park, "Transparent pressure sensor with high linearity over a wide pressure range for 3D touch screen applications," *ACS Appl. Mater. Interfaces*, vol. 12, no. 14, pp. 16691–16699, Apr. 2020, doi: 10.1021/acsami.0c00267.
[36] R. Kawakatsu and S. Hirai, "Rubbinput: An interaction technique for wet environments utilizing squeak sounds caused by finger-rubbing," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2018, pp. 512–517.
[37] Y. Iravantchi, Y. Zhao, K. Kin, and A. P. Sample, "SAWSense: Using surface acoustic waves for surface-bound event recognition," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*. NY, USA: Association for Computing Machinery, Apr. 2023, pp. 1–18, doi: 10.1145/3544548.3580991.

[38] M. Schmitz, M. Khalilbeigi, M. Balwierz, R. Lissermann, M. Mühlhäuser, and J. Steimle, "Capricate: A fabrication pipeline to design and 3D print capacitive touch sensors for interactive objects," in *Proc. 28th Annu. ACM Symp. User Interface Softw. Technol. (UIST)*. New York, NY, USA: Association for Computing Machinery, Nov. 2015, pp. 253–258, doi: 10.1145/2807442.2807503.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

MAKOTO YOSHIDA received the B.E. degree from the Osaka Institute of Technology, Osaka, Japan, in 1998. He is currently pursuing the Ph.D. degree with the Graduate School of Science and Technology, Nara Institute of Science and Technology (NAIST). His research interests include ubiquitous computing, smart home, activity recognition, and sensing technology. He is a Student Member of IPSJ.

TOMOKAZU MATSUI (Member, IEEE) received the B.E. degree from the National Institute of Technology, Nara College, in 2019, and the M.E. and Ph.D. degrees from the Nara Institute of Science and Technology (NAIST), Nara, Japan, in 2021 and 2023, respectively. Currently, he is an Assistant Professor with the Graduate School of Science and Technology, NAIST. His research interests include ubiquitous computing, smart home, activity recognition, and elderly monitoring. He is a member of IPSJ.

TOKIMUNE ISHIYAMA received the B.E. degree from the National Institute of Technology, Nara Collage, Nara, Japan, in 2021, and the M.E. degree from the Nara Institute of Science and Technology (NAIST), Nara, in 2023. His research interests include activity recognition, machine learning, ADL sensing, and smart home.

MANATO FUJIMOTO (Member, IEEE) received the B.E., M.E., and Ph.D. degrees from Kansai University, Osaka, Japan, in 2009, 2011, and 2015, respectively. From April 2015 to September 2021, he was an Assistant Professor with the Graduate School of Science and Technology, Nara Institute of Science and Technology. From October 2021 to March 2022, he was an Associate Professor with the Graduate School of Engineering, Osaka City University. He is currently an Associate Professor with the Graduate School of Informatics, Osaka Metropolitan University. His research interests include ubiquitous computing, wireless networks, sensing technology, and elderly monitoring. He is a member of ACM, IEICE, and IPSJ. He has received several academic awards, including the IPSJ/ACM Award for Early Career Contributions to Global Research, in 2023.

HIROHIKO SUWA (Member, IEEE) received the B.E. and M.E. degrees from Gunma University, in 1998 and 2000, respectively, and the Ph.D. degree from the University of Electro-Communications, Japan, in 2006. He was an Assistant Professor with the University of Electro-Communications, from 2006 to 2014, and the Nara Institute of Science and Technology (NAIST), from 2014 to 2018, where he has been an Associate Professor, since 2019. His research interests include social informatics, real data mining, and ubiquitous computing systems. He is a member of IEICE and IPSJ.

KEIICHI YASUMOTO (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in information and computer sciences from Osaka University, Osaka, Japan, in 1991, 1993, and 1996, respectively. He is currently a Professor with the Graduate School of Science and Technology, Nara Institute of Science and Technology. His research interests include distributed systems, mobile computing, and ubiquitous computing. He is a member of ACM, IPSJ, and IEICE.