Smatable: A System to Transform Furniture into Interface using Vibration Sensor

Makoto Yoshida[1], Tomokazu Matsui[1], Tokimune Ishiyama[1], Manato Fujimoto[2], Hirohiko Suwa[1], Keiichi Yasumoto[1]
[1] Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan
E-mail: {yoshida.makoto.yn3, matsui.tomokazu.mo4, ishiyama.tokimune.io3, h-suwa, yasumoto}@is.naist.jp
[2] Osaka Metropolitan University, Osaka, Osaka 558-8585, Japan
E-mail: manato@omu.ac.jp

***Abstract*—Recently, with the spread of smart houses, the smartness of housing equipment and home appliances has progressed, and the functionality and usability of interfaces between people and equipment and between people and home appliances have become important factors. Currently, the main interfaces are remote controls, smartphone applications, and even voice recognition. Furthermore, research is also being conducted on interfaces that can be operated without having the device at hand using cameras and radio waves. However, special equipment must be installed for operation, and compatibility with room design has become an issue. In this research, we proposed a system to transform existing furniture into an interface rather than providing a new interface. The proposed system focused on vibration sensors that are small, inexpensive, and can be attached to existing furniture or hidden from view. To evaluate the proposed system, an experiment was conducted to transform existing furniture into an interface for swiping by simply attaching the vibration sensor to the existing furniture. Specifically, the system attaches four vibration sensors with synchronized output signals to a table and uses a CNN to learn the vibration data obtained from the sensors to predict the direction of the swipe. As a result, when the table and person swiping were fixed, the system could predict the swipe with an accuracy of over 0.86.**

***Index Terms*—Interfaces, Vibration sensors, Deep learning**

I. Introduction

Recently, there has been an active trend toward the use of smart home equipment and home appliances. A factor that determines the usability of smart home appliances is the "interface between humans and smart home appliances." In particular, the interface of smart home appliances used in daily life must be easy to use, simple to use, fun to use, and improve the quality of life. The interaction provided by these interfaces is a key elemental technology that determines whether smart appliances will be integrated into our daily lives. Currently, the mainstream interfaces between smart home appliances and humans include remote controls, portable devices such as smartphone applications, and voice recognition. Although voice recognition has the advantage of being non-contact, it is command-based, with one command for each action, and when repeated operations are performed, such as turning pages, the user must also repeat the pronunciation, which can be stressful. In addition, efforts are being made to embed highly functioned interfaces in home appliances themselves, and products such as electrostatic touch sensors and touch panels that function as interfaces are being developed. Embedding interface functions have the advantage of being intuitive and natural. However, home appliances and furniture are specially designed for this purpose, and the scale becomes large, resulting in costs other than for the original function and design. On the other hand, non-contact interfaces are not only systems that use voice recognition, but also those that use gestures such as fingers and arms. For example, there is an interface that uses a motion sensor to operate the CG on the screen when arranging three-dimensional objects in the initial stage of work, such as design, although it is not for domestic use [1].

This method requires a grasping motion in space, so a monitor is necessary, and the motion sensor must be installed so that the hand can be seen, which limits the design of furniture when applied in the home. Additionally, there are camera-based systems that use special gloves to detect human hand movements [2], and systems that use an infrared light on the palm to capture images with an infrared camera to recognize arm and hand movements and input data [3]. Moreover, a camera that captures the movement of a hand on a GUI projected by a projector and inputs commands have been studied [4], [5].

Since cameras can easily obtain a large amount of information and identify moving objects, they can be used not only for gesture recognition as an interface but also for various applications such as human behavior recognition and anomaly detection in the home. However, the amount of information handled is large, and there are concerns about privacy invasion, especially in the home. Therefore, the camera's presence in residence causes psychological stress. Recently, methods using radio waves and vibration sensors have been proposed as moving object recognition methods less likely to infringe on privacy than cameras [6]–[14]. Table 1 shows a comparison of various sensors that can detect human movements. Vibration sensors, in particular, have the advantage of being easy to install, cost-effective, and have little concern about the infringement of privacy. Additionally, it is characterized by its concealability, which allows the sensor to be placed in an invisible position, and is a great advantage when considering its use as an interface for devices used in the home, where design is important.

In this research, we propose a method that uses this feature

TABLE I
SENSING METHOD COMPARISON TABLE

	Ease of privacy considerations	Small installation size	Abundance of information	Lo light	Sensor covertness	cost
Camera	×	×	✓	×	×	×
Microphone	×	✓	-	✓	-	✓
Laser	-	×	-	✓	×	×
Infrared	✓	×	-	✓	-	×
Radio wave	✓	×	×	✓	×	-
Vibration	✓	✓	-	✓	✓	✓

of vibration sensors to connect smart home appliances and users simply by simply attaching sensors without changing the design of existing furniture and use it as a furniture interface that blends into our lives. In this paper, we took a table as one piece of furniture and examined the detection of swiping motions on the table using a vibration sensor as a basic function for making it an interface. Section II of this paper introduces existing studies and issues related to this research. Section III describes the proposed system configuration, Section IV discusses the signals output from the vibration sensor, Section V describes the swipe identification method using CNN, its evaluation results and discussion, and finally, Section VI provides a summary and future work.

II. RELATED WORK

This section introduces several existing studies related to this research, summarizes their problems, and describes the issues to be addressed in this research and the proposed methodology.

A. Tangible user interface

Patten *et al*. [13] studied and proposed a system called "Tangible User Interface." This system uses an object on a table to which a knob called a sensor puck is attached. The object is equipped with a coil, which is used to make an electromagnetic coupling with the table to track the object's state and position. In addition, an image reflecting the state and position of the knob is projected onto the table by a projector, allowing for simultaneous input and feedback from the user. This research is excellent in that it recognizes the user's manipulation of the object and reflects that feedback as a projected image, thus providing a direct manipulation of the object and a more intuitive interface. However, it requires dedicated equipment for both the sensor pack on the table and the table itself. Therefore, it is not possible to use the interface in the way we want to achieve it, which is to simply attach a simple sensor to an existing piece of furniture.

B. Research on gesture recognition using camera indoors

In recent years, some research has been conducted using camera-based gesture recognition as an interface. Goto *et al*. [4] used a camera with a pan-tilt function and a projector to project GUI images on a table, wall, or other desired location in the home. By implementing a combination of background subtraction and skin color extraction from video, hand gestures are detected as if they were touch panels. Simone *et al*. [5] have constructed an interface by combining a camera, a projector, and an infrared laser. This interface not only recognizes human actions on the projected image but also performs object recognition. In the example of the use case of cooking support, the projector projects a cooking recipe and also realizes recognition of food ingredients. In addition, multi-touch is also realized using infrared lasers. However, these methods are not only large in scale but also require the installation of a camera in a position where the projected surface and hands are within the angle of view, and nothing obstructs the camera so that the camera can recognize the projected GUI and hands, which limits the positioning relationship with furniture and other objects. In addition, the camera-based method not only raises privacy concerns by installing a camera inside a room but also creates psychological concerns even when the camera is not in use, since the camera is installed in a position where it is easily visible to the user due to the angle of view. Furthermore, from a practical standpoint, the amount of light stored in the image sensor is reduced in many cases where the lighting in the home is darkened, resulting in blurred images and difficulty in increasing the frame rate, which increases the difficulty of identifying moving objects. In terms of operation, the amount of information handled at the time of data acquisition is large, making pre-processing computationally expensive.

C. Recognition of handwritten input using acoustic features

Other interfaces use acoustic signals to detect human writing actions. There is research on a pen interface that uses a microphone attached to the body of a pen to pick up acoustic signals generated on a desk during writing and recognize the characters in printed type [15], and research on a pen interface that uses a microphone to record writing sounds generated near a table from a distance, and combines them with character and language models estimated from the writing sound data [16], and others. These methods are superior in terms of cost and have a small installation scale, but they are easily affected by ambient noise, and there is a concern that the recognition rate may decrease under noise conditions.

D. Gesture recognition using electromagnetic waves

Kim *et al*. [14] used millimeter waves to perform gesture recognition by machine learning using the amplitude variation of the peak component of the impulse response and decision trees. This method is superior in that it does not take privacy information. However, it requires the installation of radio wave absorbers and horn antennas to suppress the effects of reflected electromagnetic waves, which is large-scale. In addition, there is concern that the use of metal furniture may change the propagation of radio waves.

E. Positioning of this study with respect to existing research issues

When considering the method of moving object recognition used as an interface in the home, infrared rays, lasers, electromagnetic waves, and vibration sensors are candidates,

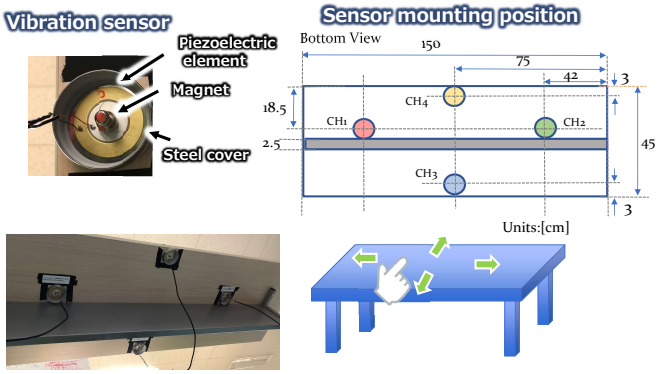


Fig. 1. Sensor Structure

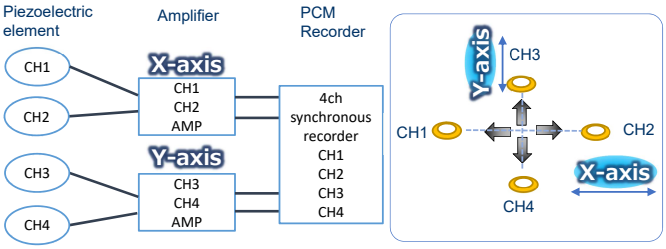


Fig. 2. block diagram

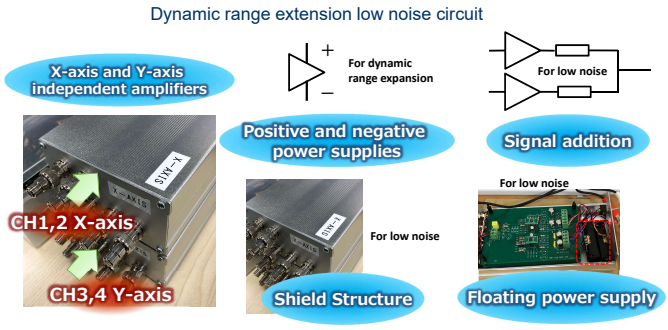


Fig. 3. Developed sensor amp system

except for cameras and microphones due to privacy concerns. However, infrared rays and lasers require sensors to be placed in a position with good visibility, which limits the degree of freedom in arranging furniture. In addition, since electromagnetic waves are reflected, furniture using electromagnetic waves is easily affected by the surrounding environment, and there is a problem that the materials used for furniture are limited. On the other hand, the vibration sensor does not provide as rich information as the camera, so privacy can be preserved. However, it has the disadvantage that it cannot obtain information when it is stationary, but it can be said that it is suitable for detecting moving objects because it contains a lot of information about moving objects. In addition, it has the advantage of being low cost and easy to arrange in large numbers.

In addition, it has the greatest feature that it can be used even when the sensor is hidden as long as it is in a place where vibration is transmitted, which is a great advantage in sensing at home where design is important. From the above, we believe that vibration sensors are suitable as interface devices for smart homes. In this research, we will develop and propose a vibration sensor that can be used as an interface with furniture just by attaching it without changing the design or material of the furniture. In this research, we used CNN, which is often used in speech recognition and image recognition. The reason is that the band of the signal recorded by the vibration generated when swiping the table is the voice band, and it was confirmed that the spectrogram clearly shows the characteristics of the swipe. The swipe detection range was set to 4 directions on 2 axes, and detection was limited to the axial direction only. Specifically, we aimed to detect four actions: swiping up (UP), swiping down (DOWN), swiping left (LEFT), and swiping right (RIGHT).

III. Sensor System

A. Sensor structure

Fig. 1 shows the structure and installation of the table and sensors used to detect swipes. Two pairs of vibration sensors are aligned in a straight line on the back of the table and mounted in orthogonal directions in the X and Y axes, respectively, for a total of four sensors to capture the vibration that occurs when a finger swipes across the table. The sensor is attached to the back of the table so that the sensor cannot be seen from the table so as not to impair the design. In previous studies, when fabricating vibration sensors, a weight of a certain mass was placed on top of the piezoelectric element to increase the sensitivity of the electrical signal generated when the object is subjected to acceleration. However, weight cannot be used when the piezoelectric element is mounted on the back side of a table. As shown in Fig. 1, the piezoelectric element is placed on a steel-covered plate attached to the table side with double-sided tape and fixed with a magnet, thus utilizing magnetic force as a substitute for the gravity obtained with a weight.

B. Sensor system block

Fig. 2 shows a block diagram of the developed sensor system. Since there are four sensors in total, each signal is amplified by a dedicated amplifier and recorded synchronously by a multi-channel recorder.

C. Sensor amplifier

Fig. 3 shows the 4-channel amplifier system fabricated. Vibration signals are known to have a very large dynamic range, i.e., the difference between strong and weak signals. The amplifiers were designed to have rail-to-rail outputs using both power supplies to ensure a dynamic range. The amplifier circuit and substrate are designed to reduce noise and ensure the signal-to-noise ratio by using parallel synthesis for the input stage amplifier.

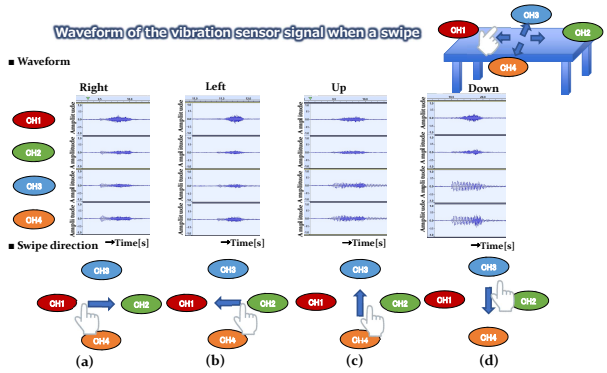


Fig. 4. Time axis data of vibration sensor output during the swipe

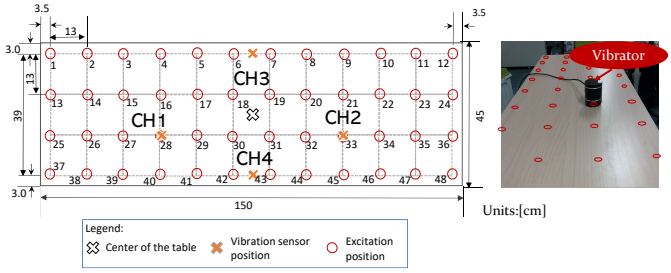


Fig. 5. Measurement environment for table vibration transfer characteristics

TABLE II
EXCITATION FREQUENCY.

Frequency[Hz]							
100	200	400	1K	2k	4k	8K	10K

In addition, the amplifier boards for the X-axis and the Y-axis are housed in the housing as a set of two channels. Additionally, it has a shield structure using an aluminum case to avoid the influence of radiation noise from electric power lines and other electronic devices when installed indoors.

Furthermore, the power supply has a floating power supply structure using batteries, and separate power supplies are installed for the X-axis and the Y-axis, thereby avoiding induced noise from power lines. These systems are for experiments and are large, but they can be made smaller by using chip parts and laminated substrates. The sample rate of data acquisition is 44.1 kHz, and quantization is performed with 16Bit PCM. The data to be recorded is recorded as one file for every two channels and recorded as two stereo WAV files, one each for CH1 and CH2 for the X-axis and CH3 and CH4 for the Y-axis.

IV. Analysis of Vibration Sensor Signal During Swiping

A. Confirmation of time domain waveform of vibration signal

Fig. 4 shows the time-axis waveforms obtained from the vibration sensor by swiping a finger on the table. The four channels of the sensor signal, CH1, CH2, CH3, and CH4, are synchronized in time. In Fig. 4, (a) is the time axis waveform when swiping to the right from the left, (b) is the time axis waveform when swiping to the left from the right, (c) is the time axis waveform when swiping down from the top, and (d) is the time axis waveform when swiping up from the bottom. Swipes were made near the center of the table.

At first, we thought that the signals of the sensor near the start position of the swipe and the sensor near the end position of the swipe would change as follows. We expected that the amplitude of the sensor near the start position would gradually decrease as the swiping finger moved, and conversely, the signal from the sensor near the end position would gradually increase. Contrary to this, we could not find a clear increase/decrease change in the amplitude of the time domain waveforms on each axis.

B. Transmission characteristics of vibration in table

To investigate the relationship between the distance from the vibration source and the output of the vibration sensor in more detail, we conducted the following experiment. As shown in Fig. 5, the table surface was divided into sections of 130 mm in length and width at equal intervals, and 48 points of intersection were vibrated by a vibrator at constant power, one at a time, with sine waves of 8 different frequencies as shown in Table 2, and the sensor output level was recorded.

For the data collected at 48 locations, the distance was calculated from the relationship between the excitation position and the sensor position, and the relationship between the distance and the sensor output level was checked. Initially, we thought that the sensor output level was inversely proportional to the distance. Fig. 6 shows the results for CH4 excitation at 400 Hz. Contrary to expectations, the amplitude level showed no inversely proportional relationship to the distance between the sensor and the exciter. This indicates that it is difficult to detect the direction of the hand at the time of swiping only by the amplitude level.

C. Auditory evaluation of vibration sensor signals

Since the recorded vibration data was recorded in PCM and its frequency range is the same as the audible range, we confirmed the vibration signal by hearing it as shown in Fig. 7. We confirmed the vibration signals by placing CH1 on the left ear and CH2 on the right ear on the X-axis, and were able to recognize them as sound and clearly feel the movement of the fingers. Similarly, when the vibration signal was checked by placing CH3 in the left ear and CH4 in the right ear on the Y-axis, the movement of the fingers could be clearly felt.

D. Confirmation by spectrogram of vibration signal

From the results of the auditory confirmation experiment in section IV-C, we thought that when the vibration signal was converted to sound, humans would be able to perceive movement as changes in pitch and tone. Therefore, in this research, instead of using the amplitude information as it is, we considered performing movement detection using the spectrogram as a feature amount. Fig. 8 shows the spectrogram calculated using FFT (Fast Fourier Transform) after recording

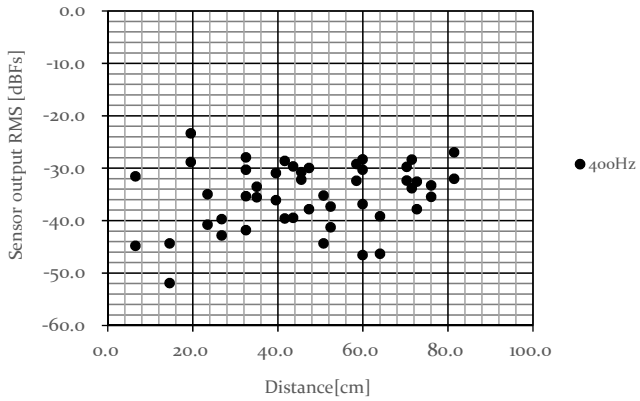


Fig. 6. Distance between vibration source and sensor VS sensor output[CH4]

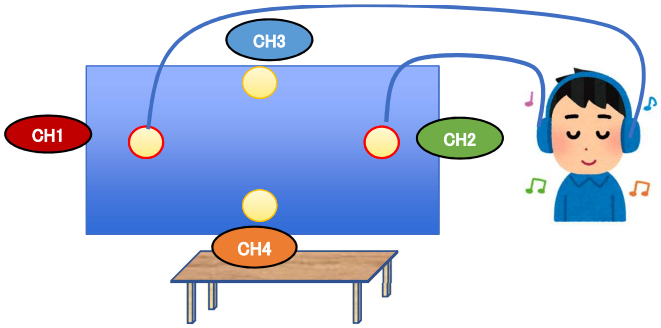


Fig. 7. Auditory evaluation of vibration sensor signals

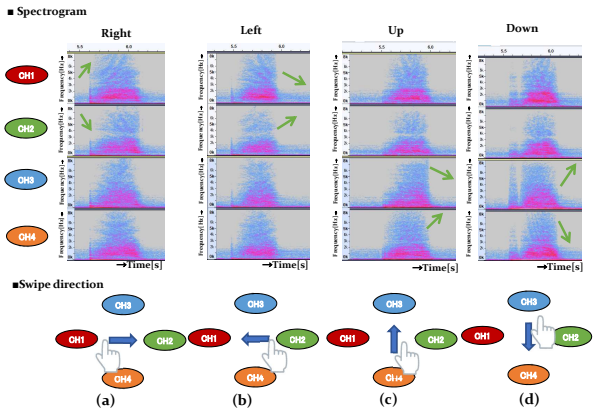


Fig. 8. Spectrogram of the signal at the swipe

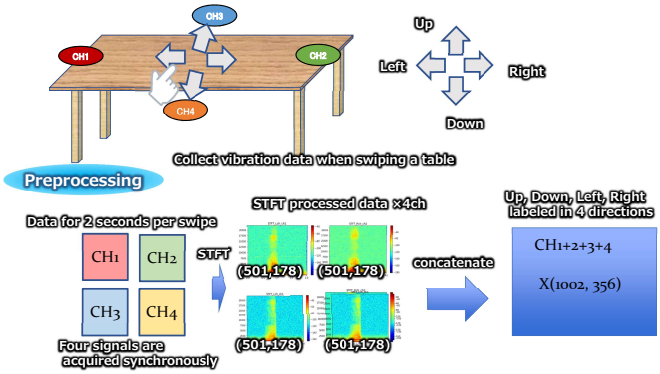


Fig. 9. Preprocessing of sensor signal data

the signals of CH1, CH2, CH3, and CH4 in synchronization. Swipes were made near the center of the table.

In Fig. 8, (a) is swiped from left to right (hereafter referred to as RIGHT), (b) is when swiped from right to left (hereafter referred to as LEFT), and (c) is swiped from top to bottom (hereinafter referred to as DOWN), and finally (d) is the spectrogram when swiping from the bottom to the top (hereinafter referred to as UP). The obtained spectrogram shows that, for horizontal swipes, the signal from the sensor at the starting point of the swipe shows a faint upward-sloping striped pattern, and the sensor at the ending point shows a faint downward-sloping striped pattern. On the other hand, for the data in the vertical direction, the change in the density of the striped pattern was small, and it was difficult to distinguish visually, but it was confirmed that a similar relationship was maintained in some data.

These stripes are thought to correspond to changes in the fundamentals and harmonics, i.e., what is perceived by the auditory sense as changes in pitch and timbre. Based on the above, we thought of using a pattern that combines four spectrograms as input and identifying the swipe direction using the temporal changes in timbre and pitch as features. Since temporal changes in pitch and timbre are expressed as changes in the pattern when the spectrogram is captured as an image, we investigated classification using CNN (Convolutional Neural Network), which is used in image recognition.

V. Swipe Direction Estimation Using CNN

A. Data acquisition and preprocessing

When we checked the signals obtained from the four sensors in section IV-D as a spectrogram, we found that a characteristic pattern appeared (Fig. 8), and furthermore, the pattern differed depending on the direction of swiping. Therefore, we decided to combine the four spectrograms into one, as shown below, and treat them as a single image-like matrix connected to the swipe direction. Fig. 9 shows the details of the preprocessing performed before deep learning. Data is cut for each swiping direction (UP, DOWN, LEFT, RIGHT) for two seconds, which contains one swipe, and converted into a spectrogram by STFT (window function is Han window, frame size is 1000 samples). Since signal data is obtained synchronously from the 4-channel sensors, the same process is performed for each of the four channels, and then the resulting four matrices for the four channels are combined as a single matrix, labeled and stored with the swiped direction (UP, DOWN, LEFT, RIGHT).

B. Acquisition of training and validation data

Data acquisition for training and validation consisted of 10 swipes per session for each swipe direction (up, down, left, right), and nine sessions of data were acquired in each direction. Swipes were made near the center of the table. In

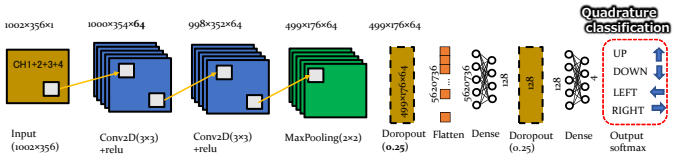


Fig. 10. Layer structure of the CNN

addition, when he checked the signal level of the obtained 2-second swipe data, it was lower than expected, so he uniformly performed a numerical amplification (multiplication) of six times (15.6 dB).

C. CNN layer configuration

As mentioned above, recognizing patterns in a spectrogram grouped into a single matrix is similar to image recognition, so in this research, we used CNN, which is often used in image recognition, for deep learning. Fig. 10 shows the layer structure of the CNN used. The obtained matrix is used as an input and trained by a CNN consisting of 2 convolution layers with 64 filters, a pooling layer, 2 dropout layers, and 2 fully connected layers. This configuration was created with reference to the configuration used for character discrimination. Specifically, the array size of the input layer is an array of (1002, 356) obtained by synthesizing four channels into one two-dimensional array using the STFT result of one channel.

Therefore, the number of arrays was adjusted accordingly. The convolution size of the subsequent convolution layer is (3 × 3), and the activation function is ReLU. The number of convolution filters is 64, and two convolution layers are connected. The pooling layer provided after the convolutional layer has a (2 × 2) configuration with 64 layers. After that, one dropout layer with a dropout ratio of 0.25 is added, followed by flattening through a fully connected layer, again with a dropout ratio of 0.25, followed by flattening through a fully connected layer, and finally four outputs corresponding to the directions (UP, DOWN, LEFT, RIGHT) with softmax functions outputs.

D. Learning methods

As shown in Fig. 11, 6 of the 9 sessions of swipe data with 10 swipes per session obtained in section 5.2 were used for training, and the remaining 3 sessions were used as validation data to perform 3-Fold cross-validation and confirm recognition accuracy. In the learning using CNN, the batch size was set to 10, and the learning model was constructed by performing 24 epochs of learning. In constructing the learning model, the loss function is set to "categorical crossentropy" and the optimization algorithm "adm" is used. Fig. 12 shows the learning curve and loss curve for one of the 3Folds, as well as the confusion matrix resulting from predicting a 4-way swipe of the data for validation using the resulting training model.

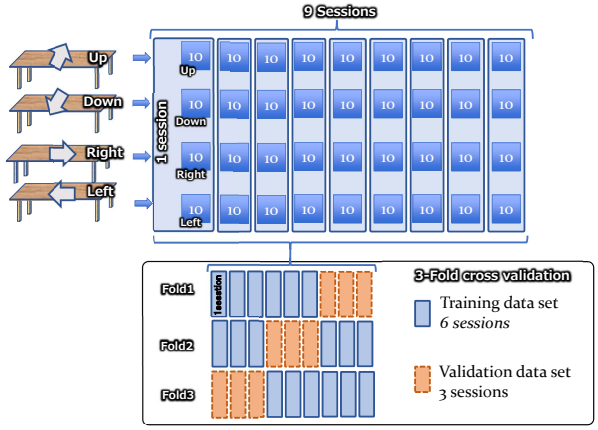


Fig. 11. 3-fold cross-validation using 9 sessions of data

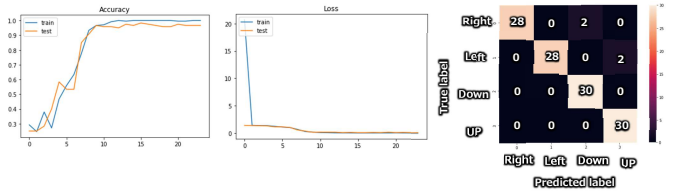


Fig. 12. An example of a learning curve and loss curve and a confusion matrix

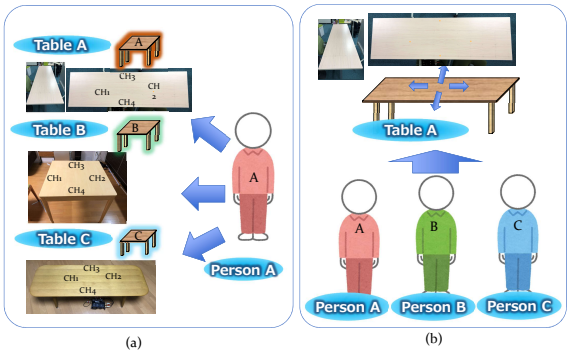
E. Comparison by person and by table

Next, in order to confirm whether the swipe direction can be identified even with different tables, as shown in Fig. 13(a), we prepared two additional tables and conducted a similar experiment. The table on which the experiment was first performed was A, and the added tables were B and C, respectively. Of the two tables, Table C chose a low table that was completely different from the other two. All the people performing the swipe motion were the same in the experiment, and the accuracy was confirmed by 3-fold cross-validation as in the experiment described earlier. The results are shown in Table 3. In all cases, the accuracy was higher than 0.9, although there was a slight decrease in accuracy in some cases when the table was changed.

Next, as shown in Fig 13(b), verification was performed when different people swiped the same table. As in the previous experiment, the results confirmed for each person by three-part cross-validation showed that the lowest accuracy was 0.86.

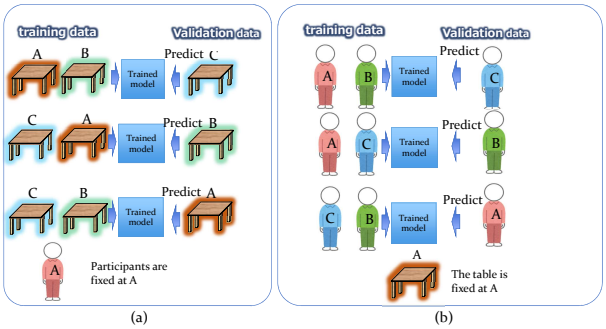
F. Accuracy in models trained on other tables and participants

As shown in Fig. 14 (a), we created a learning model using swipe data from two of the three tables that were not subject to estimation, and conducted experiments to estimate the swipe direction for the remaining one table. Fig. 14 (b), shows the results of an experiment in which a learning model is created using data from two non-targets and tested with the remaining



(a) (b)

Fig. 13. Evaluation using three types of tables and three participants



(a) (b)

Fig. 14. Evaluation of learning models using data from different tables and participants

TABLE III
ACCURACY AND LOSS WHEN USING A PER-PERSON, PER-TABLE LEARNING MODEL(3-FOLD CROSS-VALIDATION)

	Accuracy	Loss	Fixed conditions
Table A	0.98	0.05	Person A
Table B	0.94	0.32	Person A
Table C	0.90	0.60	Person A
Person A	0.98	0.04	Table A
Person B	0.86	0.61	Table A
Person C	0.92	0.47	Table A

TABLE IV
ACCURACY AND LOSS OF THE TRAINING MODEL USING DATA EXCLUDING TABLES AND PERSONS TO BE PREDICTED

	Accuracy	loss	Training data
Table A	0.56	1.30	Table B,C
Table B	0.25	2.59	Table A,C
Table C	0.22	5.03	Table B,A
Person A	0.50	2.89	Person B,C
Person B	0.39	3.95	Person A,C
Person C	0.67	0.88	Person A,B

one. The result is as shown in Table 4, and the result is a large drop in accuracy.

G. Verification of accuracy by adding data for one session of the relevant table and participants

In order to consider how to improve the accuracy, we created a learning model by adding the data from the table that tries to predict the swipe direction for one session to the learning data as shown in Fig. 15(a), and predicted the swipe direction. This is a study with a view to making it practically usable as an interface by adding a little unknown data.

Additionally, as shown in Fig. 15(b), we added the data of the person whose swipe direction is to be guessed for one session to the learning data, created the learning model again, and then Predicted the swipe direction. Results are shown in Table 5. In a per-participant experiment, all participants improved to an accuracy of 0.70 or better. In the table-by-table experiment, Table C, which is a very different type of table, showed only limited improvement in accuracy, but all other tables improved to 0.71 or better.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, on a table, which is a familiar piece of furniture in the home, the identification of human swipe operations was performed using only the vibration transmitted to the table without using a camera. The system can be implemented simply by attaching a simple vibration sensor, and there is a possibility that it can be used as an instant interface anywhere, not just on the table.

The prediction was performed using a CNN, inputting changes in the vibration signals received from four of his proprietary synchronized synchronous vibration sensors as combinations of patterns obtained from the spectrogram. In the experiment, three persons each swiped a table for multiple sessions, and the table vibration was recorded using four vibration sensors. The recorded data was used to determine the direction of the swipe, and when the vibration data of only the person who swiped the table was used for 9 sessions and confirmed by 3-fold cross-validation, the classification accuracy was higher than 0.86

Additionally, we created a learning model using 6 sessions of swipe data, 3 sessions each for the other 2 persons, excluding the person who swiped, and determined the swipe direction for the 3 sessions, but the accuracy decreased to less than 0.39. When data from the person who swiped was added for one session, the accuracy improved, and predicts were made with an accuracy of more than 0.71.

Similarly, We created training data using 6 sessions out of 9 sessions of vibration data on the table to be predicted and confirmed the prediction of the swipe direction for the remaining 3 sessions of vibration data by 3-fold cross-validation. As a result, although the accuracy varied depending on the table, we were able to make predictions with high accuracy of 0.90 or more.

In addition, we created a learning model using 6 sessions of swipe data collected from two other tables, excluding the table for which we wanted to determine the swipe direction. Using this learning model, the swipe direction was determined for 3 sessions of data in the swipe direction determination target table. As a result, the accuracy decreased to 0.22 or less.

Therefore, when we added the data of the table swiped for

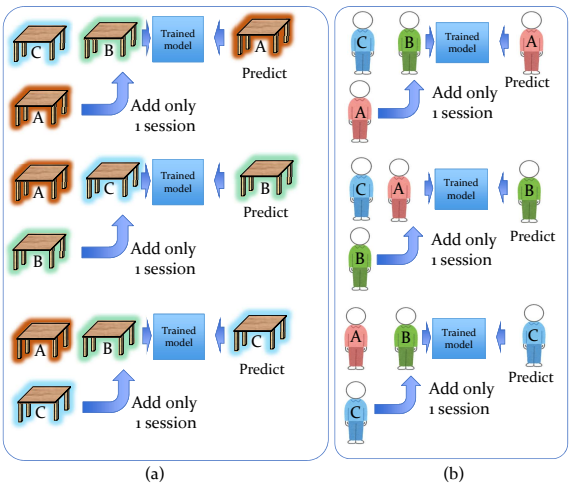


Fig. 15. Evaluation on a learning model with one additional session of data to be predicted

TABLE V
ACCURACY AND LOSS WHEN ADDED TO ONE SESSION TRAINING DATA

	Accuracy	loss	Training data
Table A	0.73	0.78	Table B,C +A(1Session)
Table B	0.90	0.36	Table A,C +B(1Session)
Table C	0.43	4.66	Table B,A +C(1Session)
Person A	0.75	1.85	Person B,C +A(1Session)
Person B	0.72	1.30	Person A,C +B(1Session)
Person C	0.76	0.73	Person A,B +C(1Session)

only one session, the accuracy improved for the two tables, excluding the table with a significantly different shape, and it became possible to estimate with an accuracy of 0.73 or more.

In this paper, we examined the possibility of using vibration data during swiping as an interface, but the swipe position was limited to the inside of the square connecting the four sensors, and the swipe was performed roughly in the middle. Therefore, it is necessary to verify the accuracy when there is a large bias in the swipe position. Additionally, at this time, the collection of training data for each table and each person operating the table is necessary for stable use, which poses a challenge.

We plan to continue the following efforts so that it can be used as an interface with only simple learning in the future. In this system, we have concluded that replaying the vibrations as sound could clearly sense movement with human hearing, so we think that improving the deep learning architecture will likely improve the detection accuracy. For this reason, we will consider acquiring generalizability in the CNN layer structure by reviewing the layer structure and architecture, such as synthesizing after performing convolution independently for each channel. For practical use as an interface, we will plan to use ensemble learning to improve accuracy and support retries in cases of low confidence.

ACKNOWLEDGEMENTS

This work was supported in part by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research number JP20H04177.

REFERENCES

[1] N. Shunsuke, K. Norihiro, “Development of motion-gesture interface to support 3d spatial modeling,” in japanese, THE 65th ANNUAL CONFERENCE OF JSSD, 2018, Vol. 65, pp. 332–333.

[2] A. Kimura, F. Shibata, T. Tsuruta, T. Sakai, M. Oniyanagi, H. Tamura, “Design and Implementation of Minority-Report-Style Gesture,” in japanese, IPSJ Journal, 47 (4),2006, pp. 1327–1339.

[3] G. Yamamoto, K Sato, “PALMbit:A Body Interface Utilizing Light Projection onto Palms,” in japanese, The Journal of The Institute of Image Information and Television Engineers, Vol. 61, No. 6, pp. 797–804, 2007.

[4] H. Goto, Y. Kawasaki and A. Nakamura, “Development of an information projection interface using a projector-camera system,” 19th International Symposium in Robot and Human Interactive Communication, 2010, pp. 50–55.

[5] S. Pizzagalli, D. Spoladore, S. Arlati, M. Sacco and L. Greci, “HIC: An interactive and ubiquitous home controller system for the smart home,” 2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH), 2018, pp. 1–6.

[6] S. Pan, T. Yu, M. Mirshekari, J. Fagert, A. Bonde, O. J. Mengshoel, H. Y. Noh, P. Zhang, “ Footprintid: Indoor pedestrian identification through ambient structural vibration sensing,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1 (3), 2017, pp. 1–31.

[7] J. Clemente, W. Song, M. Valero, F. Li and X. Liy, “Indoor Person Identification and Fall Detection through Non-intrusive Floor Seismic Sensing,” 2019 IEEE International Conference on Smart Computing (SMARTCOMP), 2019, pp. 417–424.

[8] S. Akiyama, M. Yoshida, Y. Moriyama, H. Suwa and K. Yasumoto, “Estimation of Walking Direction with Vibration Sensor based on Piezoelectric Device,” 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 2020, pp. 1–6.

[9] K. Umakoshi, T. Matsui, M. Yoshida, H. Choi, M. Fujimoto, H. Suwa, K. Yasumoto, “Non-contact person identification by piezoelectric-based gait vibration sensing, in: Proceedings of the 35th International Conference on Advanced Information Networking and Applications (2021), 2021, pp. 1–14.

[10] Y. Kashimoto, M. Fujimoto, H. Suwa, Y. Arakawa and K. Yasumoto, “Floor vibration type estimation with piezo sensor toward indoor positioning system,” 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2016, pp. 1–6, doi: 10.1109/IPIN.2016.7743667.

[11] S. Pan, N. Wang, Y. Qian, I. Velibeyoglu, H. Y. Noh, P. Zhang, “Indoor person identification through footstep induced structural vibration,” in: Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications (HotMobile’15), 2015, pp. 81–86.

[12] M. Mirshekari, J. Fagert, A. Bonde, P. Zhang, and H. Y. Noh, “Human Gait Monitoring Using Footstep-Induced Floor Vibrations Across Different Structures. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp ’18). Association for Computing Machinery, New York, NY, USA, 2018, pp. 1382–1391.

[13] J. Patten, H. Ishii, J. Hines, and G. Pangaro, “Sensetable: a wireless object tracking platform for tangible user interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’01). Association for Computing Machinery, New York, NY, USA, 2001, pp. 253–260.

[14] M. Kim,S. Moriyama, “Hand Gesture Recognition using Millimeter-wave Channel Characteristics, ” in japanese, IEICE Communications Society Conference.Vol. 1, No. BS-8-1, 2018, pp. S74–S75.

[15] M. Schrapel, M.Ludwig Stadler, and M. Rohs. 2018, “Pentelligence: Combining Pen Tip Motion and Writing Sounds for Handwritten Digit Recognition. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI ’18). Association for Computing Machinery, New York, NY, USA, Paper 131,pp. 1–11.

[16] H. Yin, A. Zhou, G. Su, B. Chen, L. Liu, and H. Ma. 2020, “Learning to Recognize Handwriting Input with Acoustic Features. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4, 2, Article 64 (June 2020), 26 pages.