Gesture Recognition Method Using Acoustic Sensing on Usual Garment

TAKASHI AMESAKA, University of Tsukuba
HIROKI WATANABE, Hokkaido University and PRESTO, Japan Science and Technology
MASANORI SUGIMOTO, Hokkaido University
BUNTAROU SHIZUKI, University of Tsukuba

In this study, we show a new gesture recognition method for clothing-based gesture input methods using active and passive acoustic sensing. Our system consists of a piezoelectric speaker and a microphone. The speaker transmits ultrasonic swept sine signals, and the microphone simultaneously records the ultrasonic signals that propagate through the garment and the rubbing sounds generated by the gestures on the garment. Our method recognizes a variety of gestures, such as pinch, twist, touch, and swipe, by incorporating active and passive acoustic sensing. An important feature of our method is that it does not require a dedicated garment or embroidery embedded since our system only requires a pair of piezoelectric elements to be attached to the usual garment with a magnet. We performed recognition experiments of 11 gestures on the forearm with four types of garments made from different materials and recognition experiments of five one-handed gestures on the button of a shirt and the pocket of pants. The results of a per-user classifier confirmed that the f-scores were 83.9% and 95.9% for 11 gestures with four different types of garments and 5 gestures that were selected assuming actual use, respectively. In addition, we confirmed that the system recognizes five gestures, which can be performed with one hand, with 89.2% and 92.6% accuracy in the button and pocket sites, respectively.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: Smart clothing, machine learning, active acoustic sensing, passive acoustic sensing

ACM Reference Format:
Takashi Amesaka, Hiroki Watanabe, Masanori Sugimoto, and Buntarou Shizuki. 2022. Gesture Recognition Method Using Acoustic Sensing on Usual Garment. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 41 (June 2022), 27 pages. https://doi.org/10.1145/3534579

1 INTRODUCTION

Wearable devices, such as hearables and smartglasses, have become more prevalent in daily life. However, these devices have the problem of operating because the user is required to touch the device itself to operate it, which limits operability. To address this problem, researchers have attempted to develop methods to extend these devices' operability by using hand and skin touch gestures. Such methods include those using electromyography sensors [14] and those using a pair of piezoelectric speakers and microphones [30, 47]. These methods require the users to directly attach the sensing devices to their skin. Hence, the user may feel uncomfortable because the devices give the user a strong sense of mechanics and wear [21]. To address this issue, *smart clothing* has

Authors' addresses: Takashi Amesaka, amesaka@iplab.cs.tsukuba.ac.jp, University of Tsukuba; Hiroki Watanabe, Hokkaido University and PRESTO, Japan Science and Technology, hiroki.watanabe@ist.hokudai.ac.jp; Masanori Sugimoto, Hokkaido University, sugi@ist.hokudai.ac.jp; Buntarou Shizuki, University of Tsukuba, shizuki@cs.tsukuba.ac.jp.

2474-9567/2022/6-ART41 $15.00
https://doi.org/10.1145/3534579

attracted attention as a future wearable sensing device; it has the potential to sense rich user information, such as gestures, posture, and biometric data, in a manner that is integrated into our daily lives [10]. Since most users in their daily lives wear clothing, the surface of the clothing could potentially be used as a touch interface, and if so, uncomfortableness can be reduced compared with the device being attached to the skin directly.

In regard to the research of input methods for smart clothing, most of them use conductive fibers to recognize gestures by detecting changes in resistance or capacitive values caused by the stretching and contact with the fabric. However, the use of conductive fibers requires a dedicated garment or embroidery embedding in the usual garment; hence, a lot of dedicated garments or embedding costs are necessary to always use the gesture input function with smart clothing since clothing has many roles, such as assisting in body temperature regulation, body protection, and self-expression, and thus different clothing is worn for different usage scenarios. Therefore, a user should be able to add smart clothing functions to a garment that they wear daily.

In this paper, we show a new gesture recognition method that uses the surface of usual garments as an input interface with active and passive acoustic sensing (Fig. 1). We focus on garment deformation and the rubbing sound generated by the gestures on the garment. Specifically, our method uses a pair of piezoelectric elements: one for transmitting ultrasonic signals and one for recording both ultrasonic signals that propagate through the taut garment and the rubbing sound during gestures. With these elements, our system can sense the deformation of the garment using an ultrasonic signal that propagates through it, and finger movements by the rubbing sound. As a result, our method recognizes a variety of gestures, such as pinching, twisting, touching, and swiping, on a garment that is difficult to propagate acoustic signals and has a changeable shape. Moreover, since our system only requires a pair of piezoelectric elements to be attached to the garment, it can be used for gesture recognition on usual garments and can be easily reconfigured the interface to other garments or sites by simply re-mounting the device.

The main contribution of this paper is the investigation of the relationship between acoustic sensing and garments (performance of each gesture, multiple garments/sites, effects of attaching/detaching the device, and effects of noise) and a deeper understanding of our system. As a result, the feasibility of our system in the most stable environment (under sitting conditions in the room), which is considered a possible usage scenario, was fully demonstrated. However, the position of this paper is a basic investigation of the feasibility of our system, not a study of its performance while walking or in many actual usage environments. In summary, the contributions of this paper are:

- We show a new gesture recognition method that uses the surface of usual garments as an input interface with active and passive acoustic sensing, which enables the user to input a variety of gestures on the garment and reconfigure the interface to other garments or sites by simply re-mounting the device.
- We implement a prototype device and investigate the effect of the acoustic properties of garments on gesture recognition through gesture experiments on the forearm using four different types of garments with 12 participants.
- We show that the average f-score of the per-user classifier of 11 gestures for four types of garments was 83.9% and that of five gestures considering actual use was 95.9%, which reveals that the user can use our system on multiple garments.
- We show that the average f-score of the per-user classifier of five gestures for the button site experiment was 89.2% and that for the pocket site experiment was 92.6%, which reveals that the user can use our system on multiple sites.

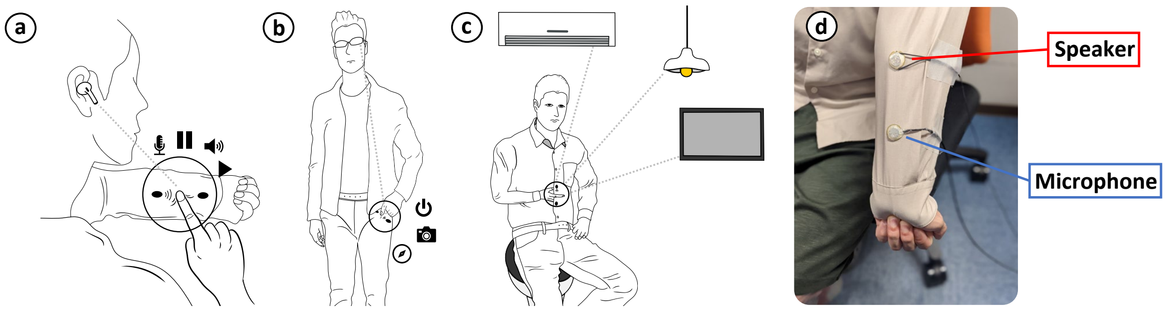


Fig. 1. Assumed scenes of system use. a: Earphone control. b: Smart glasses control. c: Smart home appliance control. d: User wearing the device. The user stretches the garment for the propagation of the ultrasonic signal during gestures. The experiment was conducted in the most stable environment where the system is expected to be used (seated state) to investigate our system's basic performance comprehensively.

2 RELATED WORK

The goal of this study is to perform gesture recognition with a usual garment surface as an interface by acoustic sensing. Our method intersects with previous studies in three main areas: smart clothing, garment-mounted devices, and acoustic sensing.

2.1 Smart Clothing

Many studies on smart clothing range from basic research to realize an embedding method with high performance in terms of implementation cost, durability, and sensing accuracy [1, 3, 9, 26, 29, 32, 34–36, 38, 39, 48] to research for practical applications.

In the research on fiber embedding methods, RESi [36] and Project Jacquard [39] are novel sensing approaches that enable a yarn-based resistive pressure sensing and capacitive touch capabilities, respectively. TexYZ [1] is a method for the rapid and effortless manufacturing of textile mutual capacitive sensors using a commodity embroidery machine. Wu et al. [48] proposed a dual-sided woven touch sensor that can recognize the interactions on the top of the bottom surface of the sensor. They also evaluated the performance of gesture recognition by embedding sensors in the cuff of a denim jacket.

Most research on practical applications focused on gesture recognition [13, 18, 37, 41, 45, 49], posture estimation [7, 8, 24, 25, 27], and vital sign sensing [17, 19, 20, 40]. In regard to gesture recognition, there are two mainstream methods: one is to use the surface of a garment as an interface to recognize touch gestures by embedding conductive fibers, and the other is to use the Doppler effect to recognize touchless gestures by placing conductive fibers to act as antennas. In the former study on touch gestures, SmartSleeve [37] is a sleeve-shaped device with a grid of conductive parts that recognizes nine gestures, including twirl, twist, fold, push, and stretch, with 89.5% accuracy. In the latter study on touchless gestures, Fabriccio [49] utilizes Doppler motion sensing with embedded conductive fibers to recognize touchless gestures, such as rub, swipe, and thumb slide. The evaluation results reported a 92.8% cross-validation accuracy and 85.2% leave-one-session-out accuracy with 11 touchless gestures and one touch gesture. In regard to posture estimation, studies have been conducted to estimate a posture by recognizing bends and twists in various parts of the body using conductive fibers embedded in clothing as stretch sensors. Corinne et al. [27] proposed a garment prototype using strain sensors to recognize upper body postures. Their evaluation showed that the recognition accuracy was 97.0% for 27 upper body postures when the classification was adapted to the individual participant. Ruibo et al. [24] proposed a system that estimates elbow joint angles by sensing the resistance change due to strain and pressure with off-the-shelf conductive fabrics and a micro-controller. The evaluation results reported that the median error was 9.69° across users with different arm sizes under motions with various speeds and magnitudes. In regard to vital sign sensing, Phyjama [20] has two types of textile-based sensors that obtain pressure changes in the textile due to cardiac and respiratory rhythm and estimate heart rate and respiration rate with high accuracy. Li et al. [17] implemented a continuous perspiration level monitoring system by sensing the amount of sweat absorption using conductive fibers covered with cotton braids. In addition to the aforementioned research on the practical application of smart clothing, Wu et al. [50] proposed a pocket-based textile sensor that detects user input and recognizes everyday objects that a user carries in the pockets of pants.

These studies using conductive fibers require a dedicated garment or embroidery to be embedded into each garment; hence, the user needs to prepare a dedicated garment and embroidery embeddings for each scenario. By contrast, our system consists of a pair of piezoelectric devices that are easy to attach and detach; therefore, our method can recognize gestures on usual garments, and the interface can be reconfigured to other clothing or sites by simply re-mounting the device.

2.2 Garment-Mounted Device

Several studies have focused on devices that can be attached and detached from a garment for gesture recognition. For example, Whack Gestures [15] is a casual eyes-free accelerometer-based device attached to a user's pocket, which can recognize a tap gesture. Zippro [22] is a zipper-shaped device with an infrared (IR) sensor, capacitive sensor, and fingerprint sensor. The research explored the possibilities of interaction with ubiquitous zipper-bearing objects, with a focus on opportunities for foreground and background interactions. SensorSnaps [6] is a button-shaped device with a low-power wireless 9-axis inertial measurement unit (IMU) sensor that can recognize tap, touch, and rotation gestures.

As with our study, the feature of these studies is that the device can be re-mounted from usual garments easily and quickly. In contrast to these studies, we focused on a gesture recognition system that uses the surface of usual garments as an interface to recognize a rich context of touch and motion gestures.

2.3 Acoustic Sensing

Acoustic sensing can be broadly classified into two types: active and passive. Research on gesture recognition and the construction of touch interfaces using each type is summarized as follows:

2.3.1 Active Acoustic Sensing. Active acoustic sensing is a sensing technique that transmits acoustic signals using speakers and captures the propagated or reflected signals using microphones. The system recognizes the state of the target by analyzing the captured signals. Watanabe et al. [47] and Kubo et al. [23] proposed a gesture recognition method using the propagation of an ultrasonic signal through a body. The Sound of Touch [30] recognizes on-body touch and hand gestures using transdermal ultrasound propagation. Takemura et al. [44] proposed a wearable sensor system that estimates the angle of an elbow and the position of a tapping finger using bone-conducted sound. These studies focus on the recognition of movement and shape of a human body by applying sound to it. In addition, several studies focus on using existing objects as touch interfaces and recognition sensors [5, 33, 43]. For example, Touch & Activate [33] is an acoustic touch sensing technique that recognizes a rich context of touches, including grasp on existing objects. SenseSurface [16] is a sensing system that can recognize an object and its position on an acrylic flat plate.

2.3.2 Passive Acoustic Sensing. Passive acoustic sensing recognizes a user's input and context by capturing and analyzing the sounds generated by a user's actions. The Sound of One Hand [2] recognizes fingertip gestures, such as tapping, rubbing, and flicking, by analyzing bone-conducted sounds. Stane [31] is a hand-held interaction device that can classify the sound generated by actions, such as scratching and rubbing. Scratch Input [11] is

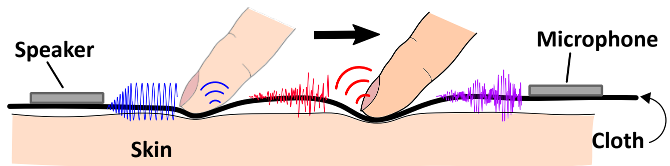


Fig. 2. Ultrasonic signal propagation and rubbing sound of garments.

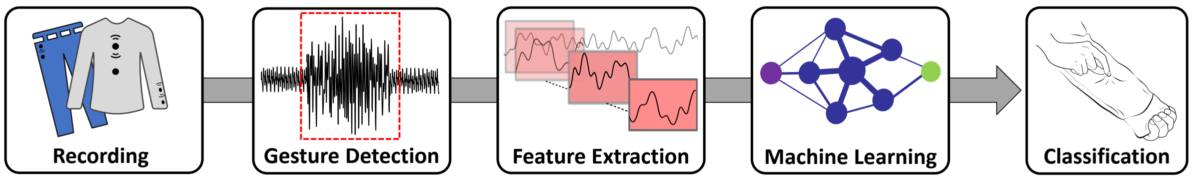


Fig. 3. Gesture recognition flow for our method.

an acoustic-based input technique using the unique sounds generated by scratching the surface of textured material. Skinput [12] estimates the tapped position on the skin by the sound propagated through the body using a microphone array. BackPat [42] identifies which finger is tapping on the back of a smartphone by analyzing the sounds generated by it. Toffee [51] recognizes the tap position around a device using acoustic time differences of arrival correlation. Mingshi et al. [4] proposed an input method that can recognize gestures and handwriting by analyzing the sound generated by sliding fingers on a table. EarBuddy [52] recognizes gestures on a surface around the ear, which was implemented in a commercial-off-the-shelf earbud.

Our method recognizes the context of touches on an object like Touch & Activate [33] and surface gestures like EarBuddy [52]. In contrast, we focused on garments as a gesture interface. Note that, clothing is difficult for acoustic signals to propagate and its shape is not as stable as the target used in previous studies [33, 52]. To address this problem, we used active and passive acoustic sensing simultaneously to recognize a variety of gestures, even for objects such as garments that are difficult to propagate ultrasonic signals.

3 OUR GESTURE RECOGNITION METHOD

In our gesture recognition method, users attach a pair of piezoelectric elements to their clothing. One is used to transmit ultrasonic swept sine signals through the garment, and the other is used as a microphone to record acoustic signals from the garment. Generally, when the garment is deformed (e.g., when the user touches the garment with a finger), ultrasonic signals obtained at the microphone change due to the cloth's deformation and absorption of the signals into the human body. When the garment is rubbed (e.g., when the user passes a finger across the garment), a rubbing sound is generated by friction (Fig. 2). Furthermore, both the ultrasonic signals and rubbing sound obtained during the gesture differ for each gesture due to the deformation of the garment, trajectory of the movement, and time that a gesture takes to finish. Hence, the frequency response changes for each gesture. With this phenomenon, our system recognizes a variety of gestures on a garment using active and passive acoustic sensing. Specifically, our system detects the occurrence of a gesture on the garment and then extracts features from the sounds of the gesture to be classified with machine learning. Fig. 3 shows the gesture recognition flow of our method. In this study, we use ultrasonic swept sine signals for active acoustic sensing as they are beyond the range of human hearing.

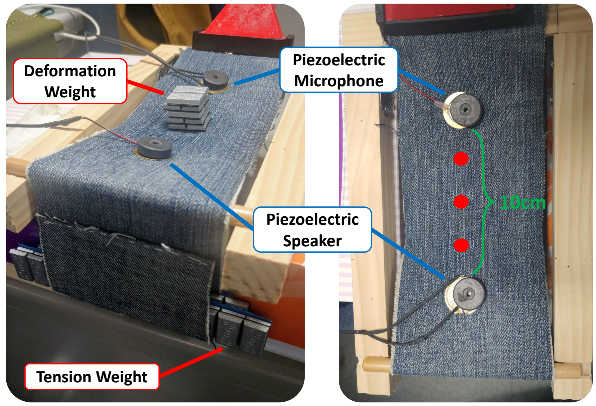


Fig. 4. Measurement environment (left) and position of the weights for deformation (right) shown as red dots.

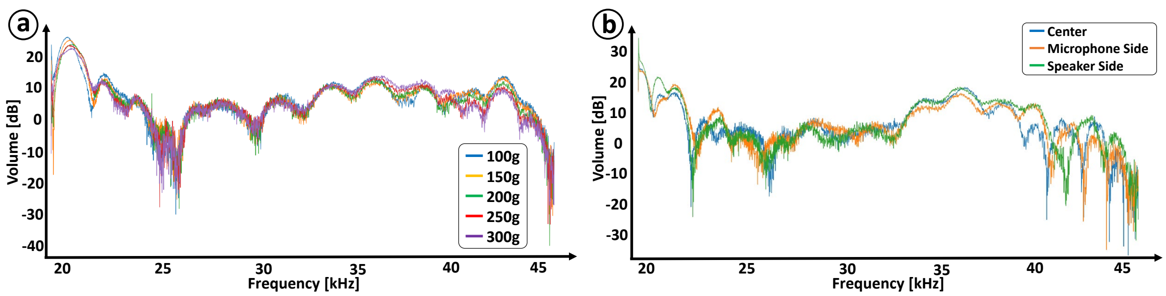


Fig. 5. Changes in frequency spectrum. a: Tension test. b: Deformation test.

3.1 Active and Passive Acoustic Sensing on the Cloth

Because we could not find any studies that used acoustic signals to recognize gestures on garments, we conducted preliminary investigations to examine whether ultrasonic signals propagating through clothing and a rubbing sound can be used for gesture recognition on the garment. To test the possibility of gesture recognition using active acoustic sensing, we observed the changes in the frequency response when tension and deformation were applied to the cloth because such changes due to tension and deformation are expected to occur during gestures. Fig. 4 shows the measurement environment. The cloth of the off-the-shelf jeans (100% cotton, non-conductive fibers, thickness: 0.50 mm) was cut into a rectangular shape and sewn so that tension weights could be placed on it. We transmitted a swept sine signal that shifted from 18 to 48 kHz over approximately 170 ms into the cloth. The sampling rate was 96 kHz, and 16-bit quantization was applied. The piezoelectric elements were a Thrive OMR20F10-BP-310 for the speaker and a Murata 7BB-20-6L0 for the microphone; they were placed 10.0 cm apart on the cloth. We used a Roland OCTA-CAPTURE audio interface to convert signals from analog to digital and digital to analog and a Lenovo ThinkPad X270 to analyze the data.

We observed the frequency spectrum in the following two tests: tension and deformation. In the tension test, we placed a 50 g deformation weight at the midpoint between the two piezoelectric elements and varied the tension weights to 100 g, 150 g, 200 g, 250 g, and 300 g. In the deformation test, we applied a 100 g tension weight and moved a 50 g deformation weight to the microphone side, the midpoint, and the speaker side. Fig. 5 shows the changes in the frequency response in the tension and deformation tests, respectively. As shown in these figures,

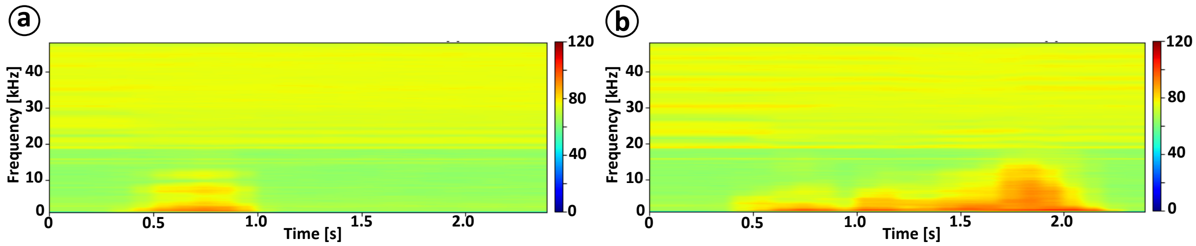


Fig. 6. Spectrograms during gestures on garment around the inside of forearm. a: Twisting. b: Drawing a circle.

changes in tension or deformation cause differences in the frequency response on a usual garment. Therefore, different gestures could generate different frequency responses since each gesture causes different changes in tension and deformation in the garment, suggesting that active acoustic sensing could be used to recognize gestures on such a garment.

Although active acoustic sensing is effective for recognizing a stationary state, such as the way of touch, by continuously analyzing the frequency response, it is ineffective at recognizing a non-stationary state, such as the trajectory trace of a swiping gesture [47]. In addition, we consider that passive acoustic sensing is effective at recognizing a non-stationary state because such gestures generate a loud sound at the interface between the object and skin. To test the possibility of gesture recognition using passive acoustic sensing, we investigated the changes in a frequency spectrogram to determine whether a rubbing sound is different depending on the gesture. Specifically, we attached a piezoelectric element on the off-the-shelf shirt (100% cotton, non-conductive fibers, thickness: 0.33 mm) around the inside of a person's forearm. Then, we recorded the sound when the same person performed different gestures around the piezoelectric element. The sampling rate was 96 kHz, and a 16-bit quantization was applied. The piezoelectric element was a Murata 7BB-20-6L0 for the microphone. Fig. 6 shows the actual frequency spectrogram when the person performed a twist gesture and drew a circle, respectively. As shown in these figures, the frequency spectrum's distribution, duration, and temporal variations were different for each gesture. This result suggests that passive acoustic sensing could be used to recognize such gestures.

3.2 Gesture Part Extraction

Before recognizing a gesture, the system needs to detect the occurrence of gestures. In this study, we detect gesture occurrences using a threshold based on a spectral moving average. By observing the spectrograms of several gestures, we found that the sound caused by the gestures was loud in the audible range at the time of contact between the garment and skin, and the spectral power remained high until the end of the gesture. Thus, we use two moving averages: short-term and long-term moving averages. We first use the short-term moving average to detect the moment of skin/garment contact and then use the long-term moving average to determine whether the gesture was performed. The moving average $MA(k)$ of the spectrogram is expressed in the following equation:

$$MA(k) = \frac{1}{n} \sum_{i=k-n+1}^{k} s_i.$$

Here, k is the k-th index of the spectrogram, n is the number of total periods, and s_i is the overall value of the i-th spectral power. We estimate the gesture occurrence using $MA(k)$ with two different values: n_s and n_l for the short-term and long-term moving averages, respectively. Let k be the start time of the gesture when the two moving average values obtained in these two total periods exceed the threshold set for each. The audio data before and after the start time of the gesture is extracted at a fixed length and used as the audio data of the gesture for machine learning.

3.3 Feature Extraction and Classification

Feature extraction was applied to the audio data of the gesture to increase machine learning efficiency. In this study, we used features that are commonly used in sound recognition since each gesture has a different completion time and spectral features, as shown in Fig. 6. Specifically, we extracted spectral features and waveform features from the audio data. For the spectral features, we used the linear frequency cepstral coefficients (LFCCs) [53]. In contrast to the Mel-frequency cepstral coefficients (MFCCs), which are used for audio and speech recognition, LFCCs use a linear filter bank to reduce the dimensions of the audio signal. We consider LFCCs to be suitable features because MFCCs do not equally extract features from the target range, considering the characteristics of human hearing. In addition to LFCCs, we used spectral features that are commonly used in sound recognition, i.e., centroid, subband peak, flux, roll-off, flatness, and bandwidth. For the waveform features, we extracted four types of features: zero-crossing rate, root mean square, variance, and attack time, which are commonly used in sound recognition. Except for the LFCCs, these features were extracted without change because they were not filtered to match the characteristics of human hearing.

In the classification, we used a support vector machine (SVM) as a classification algorithm. Although other classification algorithms, such as decision tree algorithms or ensemble learning algorithms, can be considered, we selected an SVM because the recognition result was the best. The data used for training is automatically extracted by the algorithm described in the aforementioned section; hence, the amount of data for each label is unbalanced. To make the amount of data for each label equivalent, we oversampled the other label data to match the amount of label data that was extracted the most.

4 IMPLEMENTATION

Our implementation comprises hardware, which transmits ultrasonic signals and records the sound around the garment, and software, which detects the gesture occurrence and predicts the gesture from the audio data.

4.1 Hardware

The hardware of our system consists of a pair of piezoelectric devices. The piezoelectric element on the speaker side converts electrical energy into vibrations via the inverse piezoelectric effect, generating sound. In contrast, the piezoelectric element on the microphone side converts sounds into electrical energy with the opposite effect. For efficient audio transmitting and recording, it is important that the piezoelectric elements are in close contact with the boundary surface. Hence, we adopted a method of pinching from both sides with magnets, which enables the devices to be re-mounted easily and to adhere to the garment. Fig. 7a shows the implemented device, and Fig. 7b shows the device and attaching structure. The top of the device was covered with glue so that the signal strength propagating through the garment was sufficiently larger than that propagating through the air. We observed the spectrograms in real-time and did not find Doppler shifts during gestures in the air. For signals propagating through the human body, the part that touches the skin is the magnet at the bottom, and the magnet and cloth are sandwiched between the piezoelectric element and the skin. Therefore, the signal propagating through the human body is extremely small. We confirmed that there was no change in the spectrogram, although we moved our fingertips and contracted the muscles on the inner side of the forearm while wearing the device. Fig. 8 shows the configuration of the hardware used in this study. The piezoelectric elements were a Thrive OMR20F10-BP-310 on the speaker side and a Murata 7BB-20-6L0 on the microphone side. We used a Roland OCTA-CAPTURE audio interface to convert signals from analog to digital and digital to analog and a Mouse Computer DAIV 4N to analyze the data.

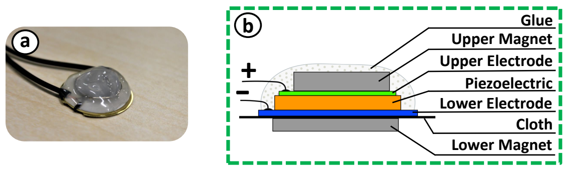


Fig. 7. Experimental devices. a: Implemented device. b: Device and attaching structure.

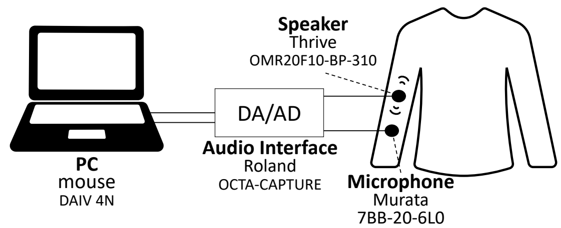


Fig. 8. Device configuration.

4.2 Software

The sampling rate for transmitting/recording sound signals was 96 kHz, and a 16-bit quantization was applied. For the active acoustic sensing, we used a swept sine signal that shifted from 18 to 48 kHz over approximately 85 ms (8,192 samples). For the frequency range, we used the maximum usable range considering inaudible for humans (18 kHz or higher) and the Nyquist frequency (48 kHz). For the signal transition time, we conducted preliminarily gesture recognition experiments with two patterns of transition time: 8,192 and 16,384 samples. However, there was no difference in accuracy; hence, we selected 8,192 samples to reduce the window size for an FFT. The speaker transmits the signal repeatedly, and the microphone records the response.

Fig. 9 shows the software flow of our system. For gesture detection and the extraction of the gesture part, we set n_s to 2 and n_l to 8. We extracted 196,608 samples (approximately $2,048$ms) for the gesture part; it begins 32,768 samples (approximately 341ms) before the start of the gesture. We use only audio data extracted using a threshold.

For feature extraction, we filtered the audio data with a 32,768-point Hamming window that includes at least three cycles of the swept sine signal. The window was shifted with 8,192 samples to extract the time variation of the features. Therefore, we acquired 24 frames ($(196,608 - 32,768)/8,192 + 1 = 24$) of sound arrays ($24 \times 32,768$) from the audio data of the gesture part. We extracted the features described in Section 3.3 from each frame; however, the attack time was extracted from the audio data of the gesture part. For the number of filter banks, we conducted preliminarily gesture recognition experiments with two patterns of filter banks: 20 and 60 banks. However, there was no difference in accuracy; hence, we selected 20 banks to reduce the number of dimensions of features. We removed the first LFCC, which is the DC component. The total number of feature dimensions obtained from the audio data of the gesture was 673.

We used a swept sine signal that shifts from 18 to 48 kHz; hence, we categorized the audio data into three frequency bandwidth patterns: *ultrasonic range* (18–48 kHz), *audible range* (0–18 kHz), and *all range* (0–48 kHz). From these frequency bandwidth patterns, we selected *audible range* and *all range* for extracting features because the recognition accuracy for the features generated by these two bandwidths had higher accuracy and was the most robust to noise. The derivation of the selected frequency pattern is described in detail in Section 5.3. The total number of dimensions of the features used for machine learning was 1,346 (673×2). We implemented the software for the measurements and data analysis using Python 3.7.

5 EVALUATION

First, we explain the experimental setup, including the selection of gestures and data collection procedures. Next, we present the evaluation of gesture extraction and the preliminary experiments for selecting frequency patterns that are important in the recognition algorithm. Subsequently, we present forearm experiments to evaluate differences in garments, the relationship between the amount of data and the recognition accuracy, and the

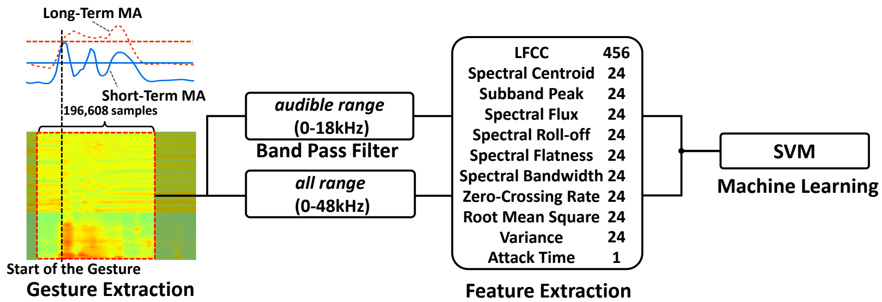


Fig. 9. Software flow of our system.

reusability of a classification model. Finally, we present experiments in the button and pocket parts to investigate gesture recognition performance for other sites.

5.1 Experimental Setup

5.1.1 Gestures. In this study, we conducted a gesture experiment using the area around the forearm as the basic input area because it is the most studied position in related works [37, 41, 45] (Fig. 10). To design the gestures used in our system, it is necessary to consider the following three conditions caused by the recognition algorithm and device configuration of our system:

- *Deformation of the garment between the speaker and microphone:* Gestures must cause the garment to deform between the speaker and microphone to change the propagation of the ultrasonic signal.
- *Asymmetrical gestures with respect to the speaker and microphone:* The movements of a gesture must be asymmetric with respect to the line connecting the speaker and microphone with other gestures. This is because two symmetric gestures, such as the upward/downward swipe where the finger crosses the line perpendicularly, could generate audio data similar to each other.
- *Gestures without touching the device:* The distance between the devices is set to 7.0 cm for the forearm experiment, which is described in Section 5.1.2. Gestures between the speaker and microphone must be performed by the fingertip because hand gestures have a high possibility of touching the element at this device's distance.

We selected six gestures (twist, touch, swipe (right/left), and pinch in/out) that met the aforementioned conditions from the related study that investigates the performance of gestures on garments [37]. In addition, we added five gestures (pick up, rotation, grasp, and circle (clockwise/anticlockwise)) that met the aforementioned conditions. As a result, we selected 11 gestures in total. A user pulls their sleeve to propagate the swept sine signal when performing a gesture. In addition, a user performs the gesture between the microphone and speaker, except for the rotation and grasp gestures. For the rotation gesture, the rotation angle was fixed at 90 degrees. For the grasp gesture, the position to grasp is on the backside where the device is located.

5.1.2 Data Collection. Table 1 shows the garments that were used in the forearm experiment. A garment consists of various parameters, such as fiber material, fabric thickness, weave, surface treatment, and fiber thickness. To investigate the effect of these differences on gesture recognition accuracy, we need to prepare the garments that differ only in one parameter. However, it would have been difficult to prepare such garments, and the combinations of the experiments would have become large. Therefore, we investigated the difference in gesture recognition accuracy following the type of garment (shirt, fleece, and jacket) in this experiment. We selected these garments because they are the outermost part of one's clothing. We collected the data for 12 participants

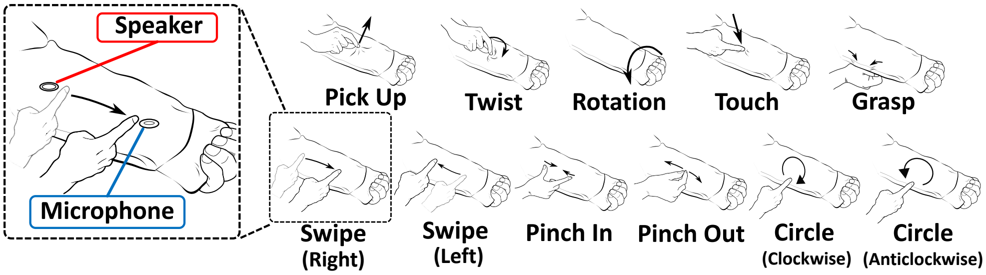
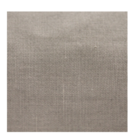


Fig. 10. Gestures around the forearm used in this study.

Table 1. Garments used in this study and their characteristics.

Appearance				
Type of garments	shirt	shirt	fleece	jacket
Label	*shirt A*	*shirt B*	*fleece*	*jacket*
Material	100% cotton	67% polyester, 33% cotton	100% polyester	100% polyester
Conductivity	No	No	No	No
Thickness [mm]	0.33	0.17	0.99	0.27
Texture				

wearing each garment. Six participants experimented on all garments. All participants were 20 to 26 years-old and were right-handed.

The experiments were conducted in our laboratory. Participants were asked to sit on a chair since recognition performance in a seated position represents the typical performance of our method under practical use because the users use seats in many cases, such as at home, in an office, and on the train. In the preliminary investigation, we observed that conversations at other desks and the sound of air conditioning did not affect the recording at the desk where the experiment was conducted. Therefore, we did not ask the other students to be quiet or to turn off any devices that generate sound, such as air conditioners. We explained the gestures and the procedure of the experiment to the participants and asked them to wear a short-sleeved top and appropriately sized garments from S to LL. Then, we attached the devices to the garment. As shown in Fig. 1d, the device was attached to the forearm of the left hand to perform gestures with the right hand. The speaker was the piezoelectric element on the elbow side, and the microphone was on the wrist side. If the distance between the devices was too short,

Table 2. Extraction accuracy of each gesture [%].

Gesture	*shirt A*	*shirt B*	*fleece*	*jacket*	Average
Pick Up	97.8	93.3	98.3	98.3	97.1
Twist	98.3	98.3	100.0	95.8	97.9
Rotation	95.0	94.2	91.7	95.0	94.0
Touch	97.2	95.8	79.2	93.3	91.5
Grasp	91.7	93.3	90.8	85.0	90.2
Swipe (Right)	98.3	96.7	95.0	98.3	97.1
Swipe (Left)	97.2	95.8	100.0	97.5	97.5
Pinch In	98.3	94.2	99.2	99.2	97.7
Pinch Out	98.9	95.8	98.3	97.5	97.5
Circle (Clockwise)	99.4	92.5	95.8	95.8	95.2
Circle (Anticlockwise)	97.8	95.0	98.3	97.5	96.9
Average	97.3	95.0	94.9	95.8	95.7

there was a high possibility that a finger could touch the devices during a gesture, which could cause false recognition from the sound generated from the collision. However, if the devices are too far away from each other, the ultrasonic signal is attenuated, and the frequency information is absent, which may decrease recognition accuracy. We determined the device distance to be approximately 7.0 cm considering a trade-off between the comfort of performing gestures and the signal attenuation. Then, we randomly showed the name and illustration of a gesture on a screen and asked the participant to pull his or her sleeve. When we observed that the participant was ready to perform the gesture, we transmitted the swept sine signal continuously, started recording, and then asked the participant to perform the gesture. We recorded the data for all gestures without removing the device and defined these recordings as one round. We asked the participant to take the device off and put it back in place between each round. In total, we acquired the audio data for 165 gestures from each participant (11 gestures $\times$ 15 rounds) in each experiment.

5.2 Gesture Extraction Accuracy

We acquired 7,920 audio data samples (12 participants $\times$ four garments $\times$ 11 gestures $\times$ 15 rounds) with the forearm experiment and automatically extracted the gestures using a threshold based on a spectral moving average. We calculated the percentage of automatically extracted data for each gesture. Table 2 shows the extraction accuracy of each garment in this study. As shown in this table, there are several gestures in which the extraction accuracy is low, such as touch and grasp. By observing the spectrogram, we concluded that the low accuracy for the touch gesture extraction was due to the short gesture time, while that for the grasp gesture extraction was due to the place where the user's fingers touched the garment being too far from the microphone. If a gesture was not extracted, although no further command will be executed, a user was required to perform the gesture again; hence, it is desirable to have high extraction accuracy from the viewpoint of usability. We need to conduct real-time gesture recognition experiments to evaluate gesture extraction and recognition simultaneously in the future.

5.3 Selection of Frequency Pattern

We consider that there are three frequency patterns for extracting features: *audible range* (0–18 kHz), *ultrasonic range* (18–48 kHz), and *all range* (0–48 kHz), since the features generated by audio data that are filtered into these ranges have different effects on machine learning; therefore, in this investigation, we compared the recognition

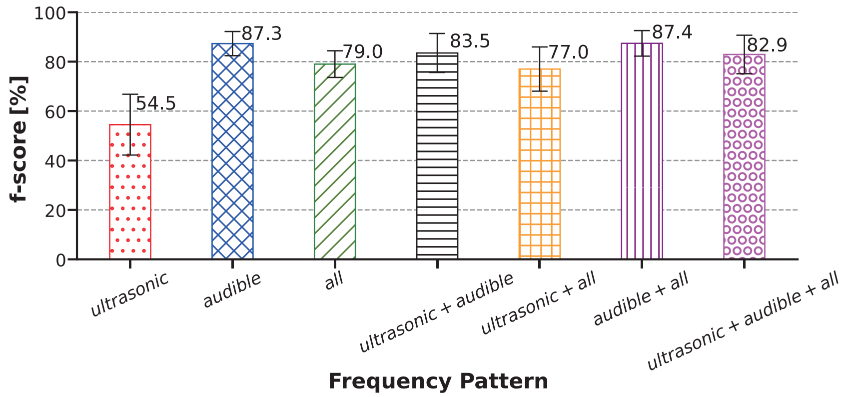


Fig. 11. Accuracy for each frequency pattern without noise. Error bars show standard f-score deviation in each participant.

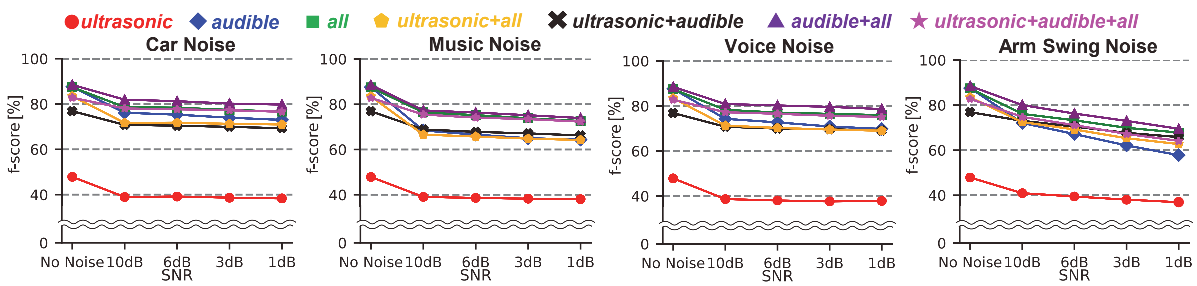


Fig. 12. Average accuracy of each frequency pattern with each noise added.

accuracies of the experiments without/with additional noise in seven possible combinations of frequency patterns: *ultrasonic range*, *audible range*, *all range*, *ultrasonic + audible ranges*, *ultrasonic + all ranges*, *audible + all ranges*, and *ultrasonic + audible + all ranges*.

In the experiment without noise, we trained the SVM classifier using the audio data of *shirt A*. Fig. 11 shows the recognition accuracy of each frequency pattern. The frequency pattern with the highest recognition accuracy was obtained by training with *audible + all ranges*. The frequency pattern with the lowest recognition accuracy was obtained by training with *ultrasonic range*.

In the experiment with noise, we prepared four different types of noise: car, music, voice, and arm swing, and added the pre-recorded noise data to the audio data using software to compare the effect of each noise at the same signal-to-noise ratio (SNR). The noise was recorded with the same microphone (Murata 7BB-20-6L0) used in the experiment. For the car noise, we recorded the running sound of a car under the condition of outdoor walking. For the music noise, we recorded the rock music from the speaker at a volume set by listening to it. For the voice noise, we recorded the voice of a user, who wore the microphone, reading some text. For the arm swing noise, we recorded the noise by swinging the arm wearing the microphone, as in the experiment. The average SNR of each noise (car, music, voice, and arm swing) was 4.9 dB, 3.6 dB, 6.6 dB, and 7.9 dB, respectively, and the average standard deviation was 3.4 dB when the obtained noise data were added as is to the audio data. These results show that the SNR distribution is roughly concentrated between 10 dB and 1 dB. Therefore, we adjusted the SNR of each noise to 10 dB, 6 dB, 3 dB, and 1 dB, respectively, to make the conditions of each noise equivalent.

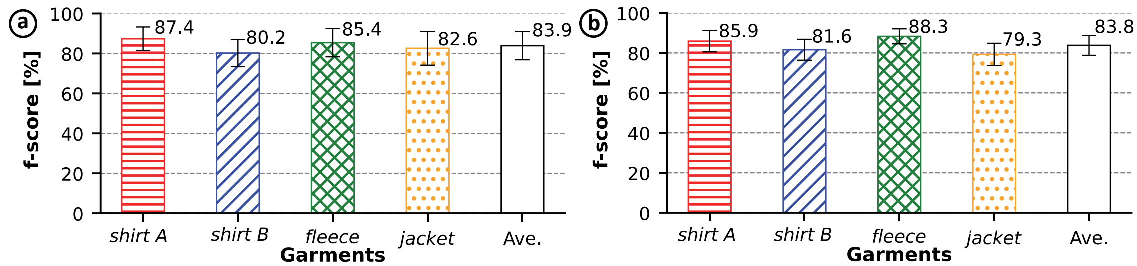


Fig. 13. Overall accuracy for each garment. Error bars show the standard deviation of the f-score in each participant. a: All participants. b: Six participants who experimented for all garments.

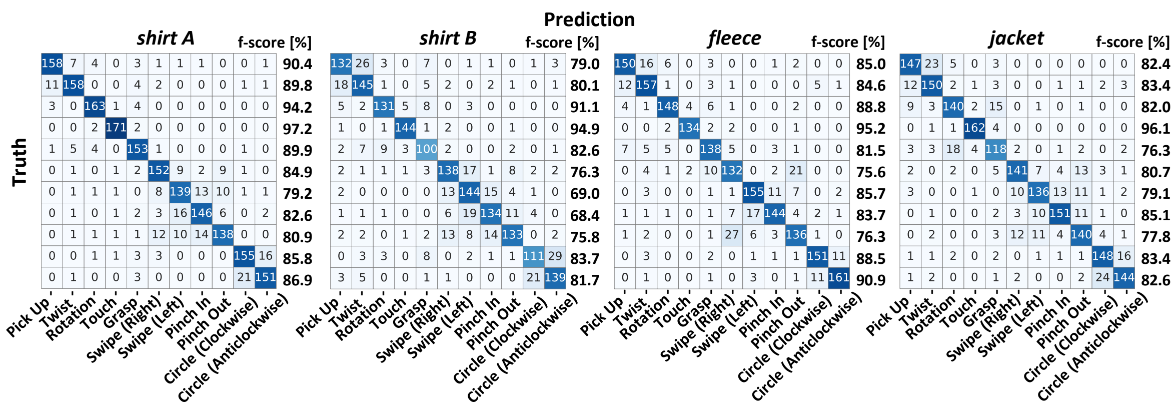


Fig. 14. Confusion matrix of all gestures with all garments.

We extracted the same features from the audio data with noise added and used these features as test data, and we trained the SVM classifier using the audio data of *shirt A*. Fig. 12 shows the change in recognition accuracy for each of the frequency patterns when the SNR was changed from 10 dB to 1 dB. For all SNRs and all noises, the recognition accuracy of the *audible + all ranges* was the most accurate. For the experiment without noise, the recognition accuracy of *audible range* was also high; by contrast, the recognition accuracy decreased significantly by 21.0 points when noise was added. The recognition accuracy of the *ultrasonic range* also decreased because the noise data contained internal noise from the electronic circuit, increasing the amount of added noise.

From the results of two experiments, we found that the *audible range* mainly contributes to the recognition accuracy. However, the recognition accuracy of *audible range*, which does not include *ultrasonic range*, decreased largely as the noise was added. The recognition accuracy of *ultrasonic range* alone was not high; however, we found that the frequency of *audible + all ranges*, including *ultrasonic range*, was robust to noise. Therefore, we concluded that the frequency pattern of *audible + all ranges* was robust to noise and had high recognition accuracy. We use *audible + all ranges* for the features in the following evaluation.

5.4 Evaluation for Different Garments

We investigated the accuracy of gesture recognition following different garment types. We collected audio data for four different garments shown in Table 1. As in Section 5.3, two types of recognition experiments were performed: data without noise and data with noise by changing the SNR.

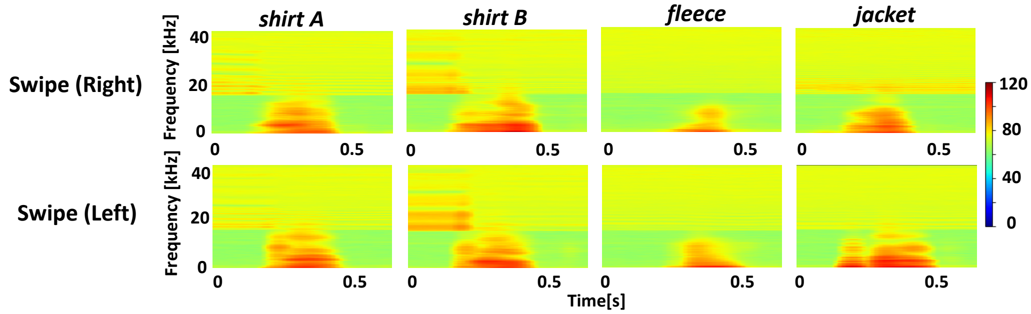


Fig. 15. Spectrograms of swipe (right) and swipe (left) for each garment.

5.4.1 Data without Noise. The overall accuracy of each garment for all participants is shown in Fig. 13a, and the average accuracy of the six participants who experimented with all garments is shown in Fig. 13b. The f-scores for *shirt A*, *shirt B*, *fleece*, and *jacket* were 87.4%, 80.2%, 85.4%, and 82.6%, respectively. Fig. 14 shows the confusion matrix of all gestures with all garments. The touch and pinch-out gestures have the highest and lowest accuracies for all garments, respectively (average f-scores 95.6% and 77.7%).

From Fig. 13, we confirmed that the recognition accuracies of *shirt A* and *fleece* are high, while those of *shirt B* and *jacket* are low. Fig. 15 shows the actual frequency spectrogram when the same person performed swipe (right/left) in different garments. As this figure shows, the spectral power obtained at the microphone decreased as the garment thickness increased. Although the spectral power of *fleece* was the smallest, the recognition accuracy of symmetrical gestures was the best among tested garments. We consider that this is because the thicker the garment is, the more sound is attenuated, and high-frequency (5–15 kHz) sound is more likely to attenuate. For example, as the spectrogram of *fleece* shows, at the beginning of the gesture of swipe right (the place away from the microphone), the microphone captured only low-frequency sound (0–5 kHz). At the end of the gesture (the place close to the microphone), the microphone captured the sound up to approximately 15 kHz. Owing to these characteristics, even symmetrical gestures have different spectrograms. The same phenomenon can be observed with swipe left. In contrast, other thinner garments (*shirt A*, *shirt B*, and *jacket*) propagated sound more easily than *fleece*; hence, the microphone could detect high-frequency sound even at the place away from the microphone. Therefore, obtained spectrograms were similar, and thus, it was more difficult to classify symmetric gestures other than *fleece*.

As a result of the aforementioned experiment without noise, the overall average of the recognition accuracy was 83.9%, which is insufficient for actual use. However, a few commands are sufficient to operate applications such as music players or comic readers, and it is expected that the recognition accuracy will be improved by narrowing down the gestures. Therefore, we will investigate the performance of the selected gestures in the following evaluation. For example, five commands are used to operate the music player: play/pause, next song, previous song, volume up, and volume down. Specifically, we selected twist, rotation, touch, swipe (right), and circle (anticlockwise) for the five selected gestures on the basis of the recognition accuracy of each gesture and the confusion performance between them. Fig. 16 shows the overall accuracy of each garment when training using only the data of these five gestures. The overall average recognition accuracy for the five gestures was 95.8%. We will conduct experiments using data from these gestures only in the following sections.

5.4.2 Data with Noise by Changing the SNR. Fig. 17 shows the change in recognition accuracy of the five selected gestures for each garment when the SNR was changed from 1 dB to 10 dB. The solid and dashed lines show the accuracies of the classifier trained with *audible + all ranges* and *audible range*, respectively. The recognition

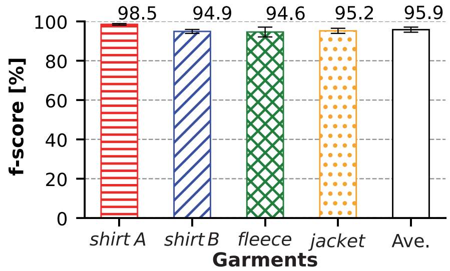


Fig. 16. Recognition accuracy of each garment and the overall accuracy considering actual use. Error bars show standard f-score deviation in each participant.

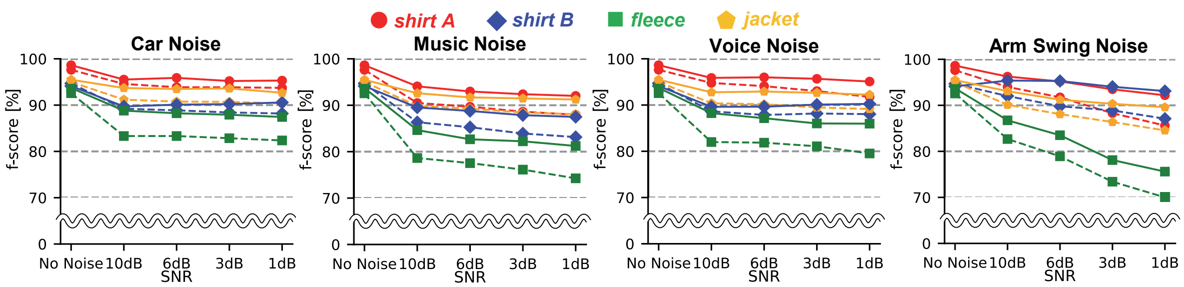


Fig. 17. Accuracy of each garment with each noise added. The solid and dashed lines show the accuracy of the classifier trained with *audible + all ranges* and *audible range*, respectively.

accuracy decreased as the SNR decreased for all garments. The recognition accuracy across the noise in each garment (*shirt A*, *shirt B*, *fleece*, and *jacket*) at an SNR of 1 dB decreased by 5.0, 3.8, 11.1, and 4.1 points from no noise, respectively. The recognition accuracy across the garment of each noise (car, music, voice, and arm swing) at an SNR of 1 dB decreased by 4.0, 7.5, 4.6, and 7.9 points from no noise, respectively. The accuracy improvement was 4.3 points at an SNR of 1 dB compared with the accuracy for the training using only data of *audible range*. The degradation of recognition accuracy was greatest for *fleece*, *shirt A*, *jacket*, and *shirt B* in that order. This order is consistent with the garment thickness (Table 1). As shown in Fig. 15, the thicker the garment, the smaller the propagation of the *ultrasonic range*. The larger the power of the *ultrasonic range*, the larger the ratio of information for the *ultrasonic range* is contained in the features for machine learning. When noise data are added to the audio data, the information in the *audible range* for the features becomes unstable; however, that of the *ultrasonic range* is stable. Therefore, the classification model generated by audio data with a large ratio of information in the *ultrasonic range* is more accurate in predicting noisy features.

5.5 Recognition Accuracy with Number of Data

In the evaluation described in Section 5.4, we used 14 rounds of data as the training data and the remaining one as unknown data to evaluate the recognition accuracy for a per-user classifier. However, to train a per-user classifier, the user must first measure the audio data. In this section, we investigate the change in recognition accuracy by varying the amount of data used for training data to evaluate the relationship between the recognition accuracy and the tolerance of the initial setup time. We conducted a questionnaire survey on the tolerance of the initial setup time. We asked participants for their impressions with a Likert scale when the time required for the initial

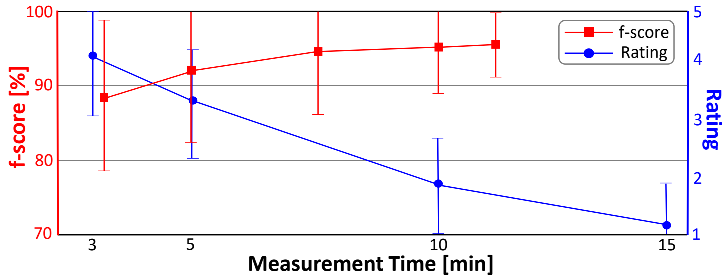


Fig. 18. Changes in f-score for the amount of training data from 4 to 14 rounds and changes in ratings of the tolerance of the initial setup time. The measurement time for one round is calculated at 50 seconds, and the amount of training data is converted to the measurement time.

setup was 3 min, 5 min, 10 min, or 15 min. The scale ranged from 1 (dissatisfaction) to 5 (satisfaction). From the questionnaire, we found that the time required for the initial setup should be less than 5 min (3 min: M=4.1, SD=1.11, 5 min: M=3.2, SD=1.15, 10 min: M=1.9, SD=0.83, 15 min: M=1.4, SD=0.60). Additionally, we trained the SVM classifier by varying the amount of data used for training data.

Fig. 18 shows the changes in recognition accuracy and standard deviation when the amount of training data is changed (red line) and the ratings of the tolerance of the initial setup time (blue line). The measurement time for the five gestures requires approximately 50 s per round (10 s $\times$ 5 gestures). This figure shows that the recognition accuracy increased as the training data increased, and the rating decreased as the measurement time increased. The recognition accuracy was more than 90% when the measurement time was greater than five min. In contrast, the rating exceeded three points within five min of measurement time; however, the rating was less than two points for more than ten min of measurement time. From the usability viewpoint, the number of required rounds should be six or less; we need to improve the recognition algorithm to obtain sufficient recognition accuracy with less than six rounds of data.

5.6 Reusability of Classification Models

To assess the reusability of the classification model, we evaluated the recognition accuracy when the gesture data of another garment of the same model or the different mounting locations were given as test data for the classification model created with the data obtained from the original garments.

5.6.1 Recognition Accuracy in Different Instance. We collected five rounds of audio data from the six participants for another garment of the same model as *shirt A*. All participants were 20 to 26 years-old and were right-handed. We then evaluated the recognition accuracy in the same manner as in Section 5.3.

Fig. 19a shows the average recognition accuracy for the six participants and the overall accuracy. For comparison, we show the accuracy of leave-one-round-out cross-validation learning using only the original data. The figure shows that the overall accuracy was 98.2% for the original data and 95.5% for another data. This recognition accuracy was approximately the same for the experiment of Section 5.4.1.

5.6.2 Recognition Accuracy in Different Mounting Location. We used gesture data at two mounting locations (wrist side and elbow side) for test data at different mounting locations. The distance between the devices was the same as in the forearm experiment. The displacement width of the device was approximately 10 cm. We collected five rounds of audio data from the six participants for each mounting location. All participants were 20 to 26 years-old and were right-handed, and *shirt A* was used. We then evaluated the recognition accuracy in the same manner as in Section 5.3.

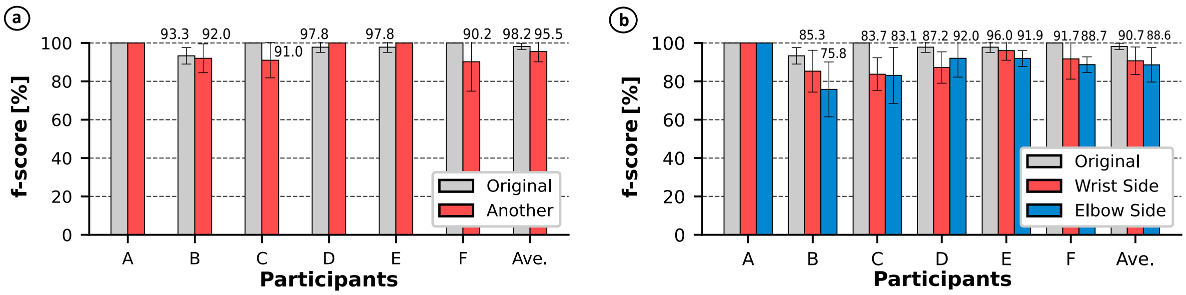


Fig. 19. Recognition accuracy of each participant and overall accuracy. Error bars show the standard f-score deviation for each participant. a: Different instance, b: Different mounting location.

Fig. 19b shows the average recognition accuracy for the six participants and the overall accuracy. The figure shows that the overall accuracies were 90.7% and 88.6% for the wrist and elbow sides, respectively. We consider that the decrease in recognition accuracy is due to the change in the garment's shape around the device caused by the change in the mounting location, which affected the acoustic characteristics. To improve the recognition accuracy, the classification model should be trained using the gesture data at each mounting location; however, this leads to a decrease in usability. To reduce the required number of training data, we consider incorporating the gesture data at each mounting location into a part of the original training data. In this experiment, we acquired data for five rounds; hence we incorporated four rounds into the original training data and evaluated the recognition accuracy using the data from the remaining one as test data. As a result of the experiment, the overall accuracies were 94.7% (*SD*=5.44) for the wrist side and 94.9% (*SD*=4.79) for the elbow side. We confirmed that the recognition accuracy was improved; however, the usability issue remains since it takes approximately three minutes to acquire data for four rounds. To avoid the need to reacquire data, we plan to investigate the changes in acoustic characteristics when the mounting location is shifted and investigate the feasibility of a robust classification algorithm for changes in mounting location.

5.7 Recognition Accuracy for General Classifiers

We evaluated the recognition accuracy of the general classifiers to examine the feasibility of our method for new users. We trained the SVM classifier for each garment using five selected gestures and performed leave-one-user-out cross-validation. Fig. 20 shows the recognition accuracy of the general classifiers for each garment. The recognition accuracy for a general classifier in each garment (*shirt A*, *shirt B*, *fleece*, and *jacket*) decreased by 7.1, 5.8, 6.3, and 5.6 points from a per-user classifier, respectively. We consider that the decrease in accuracy for the general classifier compared with the per-user classifier was due to individual differences in the speed and force of the gesture. Fig. 21 shows the distribution of each feature of *shirt A* by the t-distributed stochastic neighbor embedding (t-SNE) method [46], color-coded by the target participant and other participants. Most of the data with high recognition accuracy has a scattered distribution for each gesture (Fig. 21a). On the other hand, the data with low recognition accuracy has little variation in the distribution and shows a high degree of similarity between the participant's own data, indicating that it is not suitable for the general classifier (Fig. 21b).

In general, it is possible to use deep learning methods, such as convolutional neural networks (CNNs), to improve recognition accuracy, though, at the moment, the amount of data is insufficient, and it is difficult to improve the accuracy. Suppose our method becomes widely used in the future when the number of users increases and more data is collected. In that case, it could be possible to recognize gestures using a general classifier by CNN.

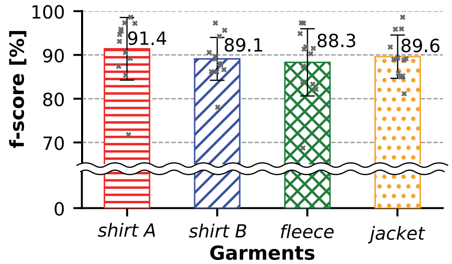


Fig. 20. Recognition accuracy of the general classifiers for each garment. Error bars show the standard f-score deviation for each participant. × is the accuracy of each participant.

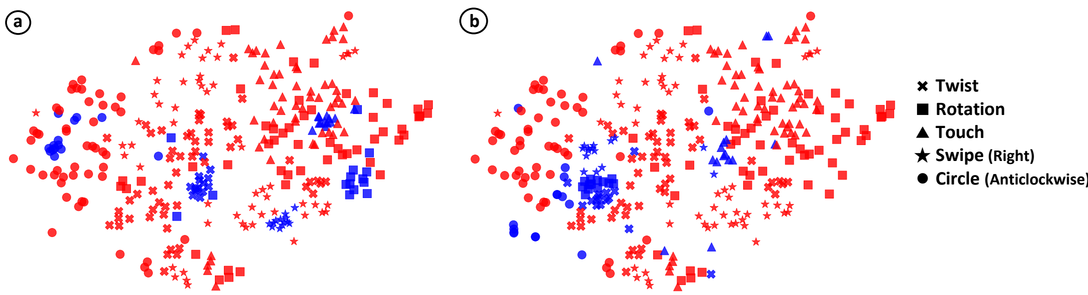


Fig. 21. Visualization of the features of *shirt A* data using the t-SNE method. Blue and red show the distributions of the target and other participants, respectively. a: The target participant with the highest recognition accuracy. b: The target participant with the lowest recognition accuracy.

5.8 Recognition Accuracy in Different Parts of Garments

To confirm that our system can be used on different sites, we conducted a gesture recognition experiment using the button of a shirt and a pocket of pants (Fig. 22a). We asked participants to pull the gesture surface by using their middle, ring, and pinky fingers to propagate the swept sine signal when performing a gesture. Fig. 22b shows the gestures for each site. The gestures were designed to be executable with one hand in consideration of usability. We also tried to minimize symmetrical gestures that can cause misrecognition. In the pocket experiment, a pair of symmetrical gestures (i.e., swipe (right/left)) was selected since the distance between the elements was too close to perform a swipe back. We collected the audio data for each site from the five participants. All participants were 20 to 26 years-old and were right-handed. In the button experiment, *shirt A* was used. In the pocket experiment, we used off-the-shelf pants (86% cotton, 9% polyester, 5% polyurethane, non-conductive fibers, thickness: 0.53 mm). We then evaluated the recognition accuracy in the same manner as in Section 5.3.

Fig. 23 shows the average recognition accuracy for the five participants and the overall accuracy. The figure shows that the overall accuracies were 89.2% for the button sites and 92.6% for the pocket sites. Fig. 24 shows the confusion matrix for each site with five participants. In the button experiment, three of the five participants had a recognition accuracy of 90% or higher, and one of them had a recognition accuracy of 95% or higher. The recognition accuracy for touch and circle gestures was higher than 90%. In the pocket experiment, three of the five participants had a recognition accuracy of 90% or higher, and two of them had a recognition accuracy of 95% or higher. The recognition accuracy for all gestures was higher than 90%.

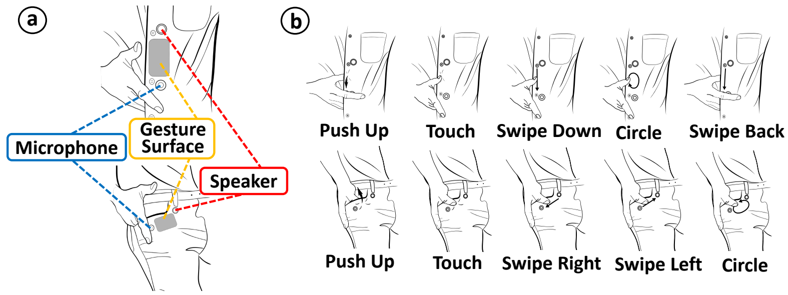


Fig. 22. a: Gesture surface and state of the hand when performing a gesture. b: Gestures used in this section.

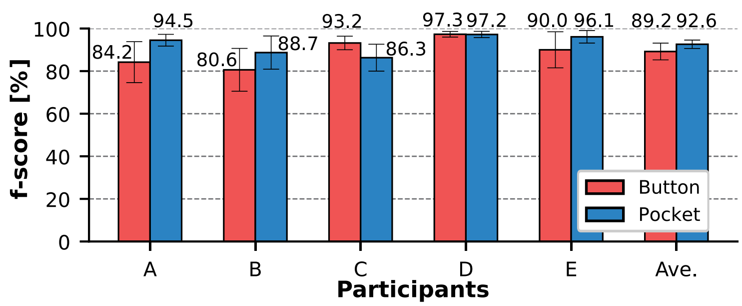


Fig. 23. Recognition accuracy of each participant and overall accuracy for each site. Error bars show the standard f-score deviation for each participant.

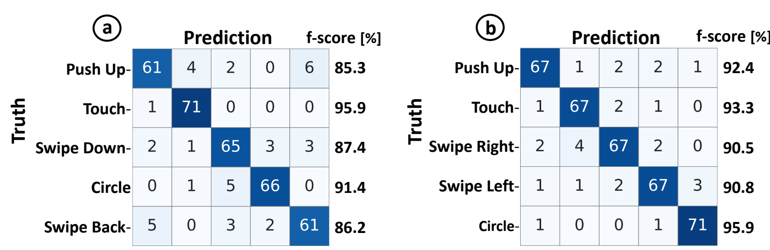


Fig. 24. Confusion matrix for five gestures with five participants. a: Button, b: Pocket.

The recognition accuracy of these experiments was lower than that in the forearm ones. This is because we could only select gestures such as tracing and touching the surface due to gesture conditions. To improve the recognition accuracy, it is necessary to select gestures with high recognition accuracy, such as twist, although it is necessary to use both hands. These results showed that the recognition accuracy of several participants exceeded 90%, suggesting that our method can be applied to other sites besides the forearm.

6 DISCUSSION AND LIMITATIONS

6.1 Recognition Accuracy of Each Gesture

There was a considerable difference in recognition accuracy between gestures. The gesture with the highest accuracy was touch (average f-score: 95.9%) for all garments because the sound generated by the touch was like a pulse, and the spectrogram obtained was very different from other gestures. The gesture with the lowest accuracy was pinch out (average f-score: 77.7%) for all garments. As shown in Fig. 14, symmetrical gestures tended to be mistaken for each other; that is, swipe (right/left), pinch in/out, and circle (clockwise/anticlockwise) were easily misrecognized. We considered that these gestures would cause a symmetrical change in the volume obtained at the microphone because the finger movements of these gestures were symmetrical; however, in several of these gestures, the volume increased even when the finger moved away from the microphone owing to changes in the force applied to the garments, which may have caused the misrecognition. There were also mistakes between swipe (right/left) and pinch in/out. We concluded that the spectrograms were similar because the trajectory and speed of the finger movements were similar for swipe (right/left) and pinch in/out. In addition, there were a number of misrecognitions between pick up and twist. These gestures have the common action of holding the garment with the index finger and thumb; thus, the obtained characteristics are similar, and misrecognition occurs.

In this study, we found that avoiding combinations of these gestures can reduce the difference in recognition accuracy for each garment (Fig. 16). To recognize these easily mistaken gestures or gestures that move symmetrically to the device (e.g., swipe upward/downward), the system could use multiple piezoelectric elements in a single device and measure the time difference of arrival to trace finger trajectories. Moreover, for the gestures that move symmetrically to the device, the recognition is still possible with the current system by using only one of these gestures. Re-examining these gestures and improving the devices are future works.

6.2 Recognition Accuracy of Each Garment

We investigated the effect of each garment on the recognition accuracy and noise in the experiments described in Section 5.4. In the experiment without noise, the results of the experiments showed that misrecognition between symmetrical gestures occurs depending on the attenuation properties of the garment. To obtain a clearer change in the spectrogram of symmetrical gestures, a longer distance between the microphone and the sound source is necessary because the attenuation of sound increases with the distance it propagates through a garment. Fig. 25 shows the spectrogram of the swipe (right/left) for *shirt B* when the device distance is varied. As shown in this figure, the longer the device distance, the clearer the change in the spectrogram of *audible range* (0 kHz–18 kHz) the larger the attenuation of the *ultrasonic range* (18 kHz–48 kHz). The attenuation in *audible range* is likely to be clearer, reducing the misrecognition of symmetrical gestures; however, the lack of information in the *ultrasonic range* may make the system less robust to noise. In the future, we plan to study the optimal device distance for each garment, considering the trade-off between recognition accuracy and robustness to noise. In contrast, although the thickness of the garment is close to that of *shirt A* and *jacket*, the obtained characteristics are different, and the recognition accuracy is higher for *shirt A* and lower for *jacket*. Thus, the difference in recognition accuracy and acoustic characteristics of the garments is not only due to the thickness of the garment but also to other factors. To further understand the gesture recognition performance of our method, a more extensive investigation of the relationship between clothing components and acoustic properties is needed.

In the experiment with noise, we found that the active acoustic sensing prevents the degradation of recognition accuracy due to noise. By contrast, Mitake et al. [28] proposed a method to prevent the degradation of recognition accuracy in noisy environments by changing the importance ratio of features obtained from acceleration and audio in accordance with the noise level. We consider that it is possible to prevent the degradation of recognition accuracy in noisy environments by changing the importance ratio of the ultrasonic range to audible range of

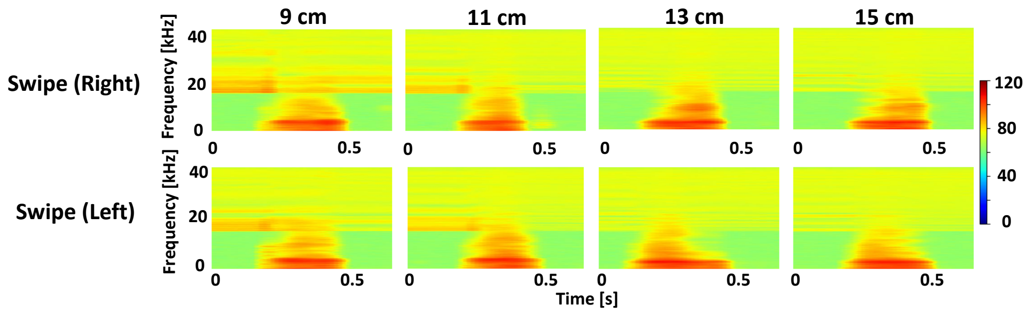


Fig. 25. Spectrograms for each distance of *shirt B* (swipe (right) and swipe (left)).

the audio data. In addition to enhancing the recognition algorithm, we plan to incorporate noise suppression methods, such as noise reduction algorithms and the soundproofing of devices.

6.3 Study of Mobile Performance

In this study, we measured the audio under sitting conditions and investigated the basic performance of our method in terms of recognition accuracy and effect of noise. Future investigations in a more practical environment are needed. We conducted a supplementary experiment with two participants for future reference. We use the data acquired in a sitting condition as training data and while walking indoors as test data. The recognition accuracy decreased from 97.8% and 95.6% to 90.2% and 64.3%, respectively, compared with the data in the sitting condition. We then manually extracted the gestures from the latter data, which had a particularly large drop in accuracy. As a result of resurveying with manual extraction data, the accuracy was improved to 77.4%. This indicates that the decrease in accuracy was caused by the gesture extraction not working well due to noise. However, the accuracy is still not sufficient. This is thought to be mainly due to motion noise and the difference in movement while walking and sitting. To address this issue, it may be necessary to use the data from walking for training. It is also necessary to investigate the recognition accuracy using audio data while standing or running, and the usability in actual applications (e.g., music player or comic reader) on a larger scale. In this study, piezoelectric devices were wired into the audio interface to conduct experiments in a stable measurement environment. We consider that it is possible to make the device mobile because our method did not require time synchronization between the speaker and microphone. We plan to make the device mobile to investigate the performance of our system in actual use.

6.4 Command Mode Detection

The system needs to switch from a non-command mode to command mode because a user will touch the garment even when there is no intention to input commands. Since our system uses a speaker and microphone, our system can use a wake word like a smart speaker or a unique sound obtained by direct touch to the microphone to switch to command mode. In addition, our system could distinguish between the non-command and command mode using sound in the *ultrasonic range* because our system requires a user to pull their sleeve to propagate the ultrasonic signal. With the command mode switch, the problem is misrecognizing the non-command mode as the command mode because an unintended command execution occurs. For this reason, it is desirable that the precision of the command mode be approximately 100%. Although the precision of the non-command mode requires high accuracy, it is less important than that of the command mode because it does not cause the system to enter the command mode.

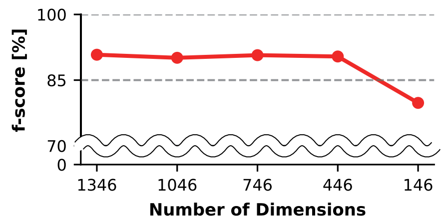


Fig. 26. Changes in f-score for the dimensionality deletion based on feature importance.

To test the aforementioned idea, we conducted a command/non-command mode identification experiment. We collected audio data for 75 s (5 s × 15 rounds) in each state, with and without pulling sleeves, for the same participants who took part in the forearm experiment. We investigated the recognition accuracy of these states for two categories: non-command mode and command mode. We extracted the data for every 32,768 samples (approximately 341 ms) and calculated the LFCCs for 19 dimensions by removing the DC component. We used an SVM in the same manner as in the previous experiment. As a result, the precision of the non-command mode was 90.2%, and that of the command mode was 72.2%. We concluded that the accuracy was not sufficient for distinguishing between the non-command mode and command mode yet. To improve the precision of the non-command mode, it is necessary to tighten up the conditions for predicting the command mode (e.g., changing the probability criterion for prediction). In future work, it is necessary to develop a more optimal method of transitioning to command mode from those methods described above.

6.5 System Limitations and Improvements

One of the limitations of our system was that it required users to execute gestures with their garments stretched; however, executing gestures at the forearm causes the problem of having both hands occupied. In the future, it is necessary to investigate the feasibility and issues of a recognition system that does not require stretching garments. In addition, there is also a possibility that the current system can be used without stretching garments by using tight-fitting garments such as sportswear. Since many users enjoy listening to music or watching videos during exercise, we believe that our system will be a good match for them. We need to carry out experiments with more types of garments to find out what kind of garments are available in our system.

We performed dimensionality deletion based on the importance of the features used in this study. As a result, the recognition accuracy was not improved by reducing the number of features and remained almost constant up to 900 deletions (Fig. 26). However, dimensionality deletion has the advantage of reducing the computational cost. In the future, the number of dimensions and features that can be deleted should be investigated in detail to reduce the computational cost.

6.6 Utility of Ultrasound

Using active acoustic sensing with ultrasonic signals, our system achieves a 0.5% accuracy improvement in noiseless environments and a 4.3% accuracy improvement in the noisy environment compared with the recognition method using only passive acoustic sensing for the five selected gestures. However, the use of active acoustic sensing requires the use of a speaker, which requires extra hardware costs. In addition, it requires a higher sampling rate than passive acoustic sensing, which increases the computational time and power consumption. One means to reduce power consumption in the current system is to change the use of active acoustic sensing in accordance with the noise level, thereby reducing the time spent using the high sampling rate. In the future, it is necessary to evaluate the utility of using ultrasound in this system after a detailed study of the cost and power consumption of such hardware.

7 CONCLUSION

We presented a gesture recognition method for usual garments using active and passive acoustic sensing. Our system recognizes a variety of gestures on a usual garment using a pair of detachable devices. We comprehensively studied the effect of differences in garments, noise type, the amount of training data on recognition accuracy, and the reusability of the classification model. The evaluation results of a per-user classifier confirmed that the f-score was 83.9% for 11 gestures with four different types of garments, and assuming actual use, the f-score was 95.9% for five selected gestures that were selected. In addition, we confirmed that the system recognizes five gestures, which can be performed with one hand, with 89.2% and 92.6% accuracy in the button and pocket sites, respectively.

ACKNOWLEDGMENTS

This work is supported by JSPS, KAKENHI Grant Number 21J10706/18H04104 and JST, PRESTO Grant Number JPMJPR2138.

REFERENCES

[1] Roland Aigner, Andreas Pointner, Thomas Preindl, Rainer Danner, and Michael Haller. 2021. TexYZ: Embroidering Enameled Wires for Three Degree-of-Freedom Mutual Capacitive Sensing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, Article 499, 12 pages. https://doi.org/10.1145/3411764.3445479

[2] Brian Amento, Will Hill, and Loren Terveen. 2002. The Sound of One Hand: A Wrist-Mounted Bio-Acoustic Fingertip Gesture Interface. In *Proceedings of the CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. Association for Computing Machinery, 724–725. https://doi.org/10.1145/506443.506566

[3] Leah Buechley, Mike Eisenberg, Jaime Catchen, and Ali Crockett. 2008. The LilyPad Arduino: Using Computational Textiles to Investigate Engagement, Aesthetics, and Diversity in Computer Science Education. In *Proceedings of the 2008 CHI Conference on Human Factors in Computing Systems (CHI '08)*. Association for Computing Machinery, 423–432. https://doi.org/10.1145/1357054.1357123

[4] Mingshi Chen, Panlong Yang, Jie Xiong, Maotian Zhang, Youngki Lee, Chaocan Xiang, and Chang Tian. 2019. Your Table Can Be an Input Panel: Acoustic-Based Device-Free Interaction Recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 21 pages. https://doi.org/10.1145/3314390

[5] Tim Collins. 2009. Active acoustic touch interface. *Electronics Letters* 45, 20 (2009), 2 pages. https://doi.org/10.1049/el.2009.2054

[6] Artem Dementyev, Tomás Vega Gálvez, and Alex Olwal. 2019. SensorSnaps: Integrating Wireless Sensor Nodes into Fabric Snap Fasteners for Textile Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. Association for Computing Machinery, 17–28.

[7] Guido Gioberto. 2014. Garment-Integrated Wearable Sensing for Knee Joint Monitoring. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program (ISWC '14 Adjunct)*. Association for Computing Machinery, 113–118. https://doi.org/10.1145/2641248.2642736

[8] Guido Gioberto, James Coughlin, Kaila Bibeau, and Lucy E. Dunne. 2013. Detecting Bends and Fabric Folds Using Stitched Sensors. In *Proceedings of the 2013 International Symposium on Wearable Computers (ISWC '13)*. Association for Computing Machinery, 53–56. https://doi.org/10.1145/2493988.2494355

[9] Nur Al-huda Hamdan, Simon Voelker, and Jan Borchers. 2018. Sketch&Stitch: Interactive Embroidery for E-Textiles. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, 1–13. https://doi.org/10.1145/3173574.3173656

[10] Alex Hanuska, Bharath Chandramohan, Laura Bellamy, Pauline Burke, Rajiv Ramanathan, and Vijay Balakrishnan. 2016. *Smart clothing market analysis*. Technical Report. University of California, Berkeley.

[11] Chris Harrison and Scott E. Hudson. 2008. Scratch Input: Creating Large, Inexpensive, Unpowered and Mobile Finger Input Surfaces. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology (UIST '08)*. Association for Computing Machinery, 205–208. https://doi.org/10.1145/1449715.1449747

[12] Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: Appropriating the Body as an Input Surface. In *Proceedings of the 2010 CHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, 453–462. https://doi.org/10.1145/1753326.1753394

[13] Florian Heller, Stefan Ivanov, Chat Wacharamanotham, and Jan Borchers. 2014. FabriTouch: Exploring Flexible Touch Input on Textiles. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers (ISWC '14)*. Association for Computing Machinery, 59–62. https://doi.org/10.1145/2634317.2634345

[14] Donny Huang, Xiaoyi Zhang, T. Scott Saponas, James Fogarty, and Shyamnath Gollakota. 2015. Leveraging Dual-Observable Input for Fine-Grained Thumb Interaction Using Forearm EMG. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology (UIST '15)*. Association for Computing Machinery, 523–528. https://doi.org/10.1145/2807442.2807506

[15] Scott E. Hudson, Chris Harrison, Beverly L. Harrison, and Anthony LaMarca. 2010. Whack Gestures: Inexact and Inattentive Interaction with Mobile Devices. In *Proceedings of the Fourth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '10)*. Association for Computing Machinery, 109–112. https://doi.org/10.1145/1709886.1709906

[16] Daiki Iwase, Yuichi Itoh, Hidehiko Shin, and Takao Onoyoe. 2019. SenseSurface: Using Active Acoustic Sensing to Detect What is Where. *Journal of Information Processing* 60, 10 (2019), 12 pages. https://ci.nii.ac.jp/naid/170000180546/

[17] Ji Jia, Chengtian Xu, Shijia Pan, Stephen Xia, Peter Wei, Hae Noh, Pei Zhang, and Xiaofan Jiang. 2018. Conductive Thread-Based Textile Sensor for Continuous Perspiration Level Monitoring. *Sensors* 18, 11 (2018), 19 pages. https://doi.org/10.3390/s18113775

[18] Thorsten Karrer, Moritz Wittenhagen, Leonhard Lichtschlag, Florian Heller, and Jan Borchers. 2011. Pinstripe: Eyes-Free Continuous Input on Interactive Clothing. In *Proceedings of the 2011 CHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, 1313–1322. https://doi.org/10.1145/1978942.1979137

[19] Ali Kiaghadi, Morgan Baima, Jeremy Gummeson, Trisha Andrew, and Deepak Ganesan. 2018. Fabric as a Sensor: Towards Unobtrusive Sensing of Human Behavior with Triboelectric Textiles *(SenSys '18)*. Association for Computing Machinery, 199–210. https://doi.org/10.1145/3274783.3274845

[20] Ali Kiaghadi, Seyedeh Zohreh Homayounfar, Jeremy Gummeson, Trisha Andrew, and Deepak Ganesan. 2019. Phyjama: Physiological Sensing via Fiber-Enhanced Pyjamas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3, Article 89 (2019), 29 pages. https://doi.org/10.1145/3351247

[21] James Knight and Chris Baber. 2005. A Tool to Assess the Comfort of Wearable Computers. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 47, 1 (2005), 15 pages. https://doi.org/10.1518/0018720053653875

[22] Pin-Sung Ku, Jun Gong, Te-Yen Wu, Yixin Wei, Yiwen Tang, Barrett Ens, and Xing-Dong Yang. 2020. Zippro: The Design and Implementation of An Interactive Zipper. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, 1–13. https://doi.org/10.1145/3313831.3376756

[23] Yuki Kubo, Yuto Koguchi, Buntarou Shizuki, Shin Takahashi, and Otmar Hilliges. 2019. AudioTouch: Minimally Invasive Sensing of Micro-Gestures via Active Bio-Acoustic Sensing. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '19)*. Association for Computing Machinery, 1–13. https://doi.org/10.1145/3338286.3340147

[24] Ruibo Liu, Qijia Shao, Siqi Wang, Christina Ru, Devin Balkcom, and Xia Zhou. 2019. Reconstructing Human Joint Motion with Computational Fabrics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1, Article 19 (2019), 26 pages. https://doi.org/10.1145/3314406

[25] Federico Lorussi, Walter Rocchia, Enzo Scilingo, Alessandro Tognetti, and Danilo de rossi. 2004. Wearable, Redundant Fabric-Based Sensor Arrays for Reconstruction of Body Segment Posture. *IEEE Sensors Journal* 4, 6 (2004), 12 pages. https://doi.org/10.1109/JSEN.2004.837498

[26] Yiyue Luo, Kui Wu, Tomás Palacios, and Wojciech Matusik. 2021. KnitUI: Fabricating Interactive and Sensing Textiles with Machine Knitting. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, Article 668, 12 pages. https://doi.org/10.1145/3411764.3445780

[27] Corninne Mattmann, Oliver Amft, Holger Harms, Gerhard Troster, and Frank Clemens. 2007. Recognizing Upper Body Postures using Textile Strain Sensors. In *Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers (ISWC '07)*. Institute of Electrical and Electronics Engineers, 29–36. https://doi.org/10.1109/ISWC.2007.4373773

[28] Hiroto Mitake, Hiroki Watanabe, and Sugimoto Masanori. 2020. A Method for Recognizing Road Surface Condition based on Footsteps and Inertial Data. *Journal of Information Processing* 61, 10 (2020), 12 pages. https://doi.org/10.20729/00207250

[29] Md. Tahmidul Islam Molla, Crystal Compton, and Lucy E. Dunne. 2018. Launderability of Surface-Insulated Cut and Sew E-Textiles. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers (ISWC '18)*. Association for Computing Machinery, 104–111. https://doi.org/10.1145/3267242.3267255

[30] Adiyan Mujibiya, Xiang Cao, Disney S. Tan, Dan Morris, Shwetak Naran Patel, and Jun Rekimoto. 2013. The Sound of Touch: On-body Touch and Gesture Sensing Based on Transdermal Ultrasound Propagation. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces (ITS '13)*. Association for Computing Machinery, 189–198. https://doi.org/10.1145/2512349.2512821

[31] Roderick Murray-Smith, John Williamson, Stephen Hughes, and Torben Quaade. 2008. Stane: Synthesized Surfaces for Tactile Input. In *Proceedings of the 2008 CHI Conference on Human Factors in Computing Systems (CHI '08)*. Association for Computing Machinery, 1299–1302. https://doi.org/10.1145/1357054.1357257

[32] Alex Olwal, Jon Moeller, Greg Priest-Dorman, Thad Starner, and Ben Carroll. 2018. I/O Braid: Scalable Touch-Sensitive Lighted Cords Using Spiraling, Repeating Sensing Textiles and Fiber Optics. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings (UIST '18 Adjunct)*. Association for Computing Machinery, 203–207. https://doi.org/10.1145/3266037.3271651

[33] Makoto Ono, Buntarou Shizuki, and Jiro Tanaka. 2013. Touch & Activate: Adding Interactivity to Existing Objects Using Active Acoustic Sensing. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. Association for Computing Machinery, 31–40. https://doi.org/10.1145/2501988.2501989

[34] Maggie Orth, Rehmi Post, and Emily Cooper. 1998. Fabric Computing Interfaces. In *Proceedings of the 1998 CHI Conference on Human Factors in Computing Systems (CHI '98)*. Association for Computing Machinery, 331–332. https://doi.org/10.1145/286498.286800

[35] Simon Ozbek, Md. Tahmidul Islam Molla, Crystal Compton, and Brad Holschuh. 2018. Novel Manufacturing of Advanced Smart Garments: Knitting with Spatially-Varying, Multi-Material Monofilament. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers (ISWC '18)*. Association for Computing Machinery, 120–127. https://doi.org/10.1145/3267242.3267278

[36] Patrick Parzer, Florian Perteneder, Kathrin Probst, Christian Rendl, Joanne Leong, Sarah Schuetz, Anita Vogl, Reinhard Schwoediauer, Martin Kaltenbrunner, Siegfried Bauer, and Michael Haller. 2018. RESi: A Highly Flexible, Pressure-Sensitive, Imperceptible Textile Interface Based on Resistive Yarns. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. Association for Computing Machinery, 745–756. https://doi.org/10.1145/3242587.3242664

[37] Patrick Parzer, Adwait Sharma, Anita Vogl, Jürgen Steimle, Alex Olwal, and Michael Haller. 2017. SmartSleeve: Real-time Sensing of Surface and Deformation Gestures on Flexible, Interactive Textiles, Using a Hybrid Gesture Detection Pipeline. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. 565–577. https://doi.org/10.1145/3126594.3126652

[38] Ernest Rehmi Post, Maggie Orth, Peter R Russo, and Neil A Gershenfeld. 2000. E-broidery: Design and fabrication of textile-based computing. *IBM Systems Journal* 39, 3&4 (2000), 21 pages. https://doi.org/10.1147/sj.393.0840

[39] Ivan Poupyrev, Nan-Wei Gong, Shiho Fukuhara, Mustafa Emre Karagozler, Carsten Schwesig, and Karen E Robinson. 2016. Project Jacquard: Interactive Digital Textiles at Scale. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, 4216–4227. https://doi.org/10.1145/2858036.2858176

[40] Emilio Sardini, Mauro Serpelloni, and Viviane Pasqui. 2015. Wireless Wearable T-Shirt for Posture Monitoring During Rehabilitation Exercises. *IEEE Transactions on Instrumentation and Measurement* 64, 2 (2015), 10 pages. https://doi.org/10.1109/TIM.2014.2343411

[41] Stefan Schneegass and Alexandra Voit. 2016. GestureSleeve: using touch sensitive fabrics for gestural input on the forearm for controlling smartwatches. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. Association for Computing Machinery, 108–115. https://doi.org/10.1145/2971763.2971797

[42] Karsten Seipp and Kate Devlin. 2014. BackPat: One-Handed off-Screen Patting Gestures. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '14)*. Association for Computing Machinery, 77–80. https://doi.org/10.1145/2628363.2628396

[43] John Lammey Stewart and E. C. Westerfield. 1959. A Theory of Active Sonar Detection. *Proceedings of the Institute of Radio Engineers* 47, 5 (1959), 10 pages. https://doi.org/10.1109/JRPROC.1959.287283

[44] Kentaro Takemura, Akihiro Ito, Jun Takamatsu, and Tsukasa Ogasawara. 2011. Active Bone-conducted Sound Sensing for Wearable Interfaces. In *Proceedings of the 24th Annual ACM Symposium Adjunct on User Interface Software and Technology (UIST '11 Adjunct)*. Association for Computing Machinery, 53–54. https://doi.org/10.1145/2046396.2046419

[45] Kentaro Ueda, Tsutomu Terada, and Masahiko Tsukamoto. 2018. Evaluation of Input/Output Interface Using Wrinkles on Clothes. *Journal of Data Intelligence* 1, 1 (2018), 19 pages. https://doi.org/10.1145/3282353.3282365

[46] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, 86 (2008), 27 pages. http://jmlr.org/papers/v9/vandermaaten08a.html

[47] Hiroki Watanabe, Tsutomu Terada, and Masahiko Tsukamoto. 2017. Gesture Recognition Method Utilizing Ultrasonic Active Acoustic Sensing. *Journal of Information Processing* 25 (2017), 10 pages. https://doi.org/10.2197/ipsjjip.25.331

[48] Tony Wu, Shiho Fukuhara, Nicholas Gillian, Kishore Sundara-Rajan, and Ivan Poupyrev. 2020. ZebraSense: A Double-Sided Textile Touch Sensor for Smart Clothing. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20)*. Association for Computing Machinery, 662–674. https://doi.org/10.1145/3379337.3415886

[49] Te-Yen Wu, Shutong Qi, Junchi Chen, MuJie Shang, Jun Gong, Teddy Seyed, and Xing-Dong Yang. 2020. Fabriccio: Touchless Gestural Input on Interactive Fabrics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, 1–14. https://doi.org/10.1145/3313831.3376681

[50] Te-Yen Wu, Zheer Xu, Xing-Dong Yang, Steve Hodges, and Teddy Seyed. 2021. Project Tasca: Enabling Touch and Contextual Interactions with a Pocket-Based Textile Sensor. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 4, 13 pages. https://doi.org/10.1145/3411764.3445712

[51] Robert Xiao, Greg Lew, James Marsanico, Divya Hariharan, Scott Hudson, and Chris Harrison. 2014. Toffee: Enabling Ad Hoc, around-Device Interaction with Acoustic Time-of-Arrival Correlation. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '14)*. Association for Computing Machinery, 67–76. https://doi.org/10.1145/2628363.2628383

[52] Xuhai Xu, Haitian Shi, Xin Yi, WenJia Liu, Yukang Yan, Yuanchun Shi, Alex Mariakakis, Jennifer Mankoff, and Anind K. Dey. 2020. EarBuddy: Enabling On-Face Interaction via Wireless Earbuds. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, 1–14. https://doi.org/10.1145/3313831.3376836

[53] Xinhui Zhou, Daniel Garcia-Romero, Ramani Duraiswami, Carol Espy-Wilson, and Shihab Shamma. 2011. Linear versus mel frequency cepstral coefficients for speaker recognition. In *Proceedings of the 2011 IEEE Workshop on Automatic Speech Recognition and Understanding*. Institute of Electrical and Electronics Engineers, 559–564. https://doi.org/10.1109/ASRU.2011.6163888