HandSAW: Wearable Hand-based Event Recognition via On-Body Surface Acoustic Waves

KAYLEE YAXUAN LI , University of Michigan, USA
YASHA IRAVANTCHI, University of Michigan, USA
YICHEN ZHU, University of Michigan, USA
HYUNMIN PARK, University of Michigan, USA
ALANSON P. SAMPLE, University of Michigan, USA

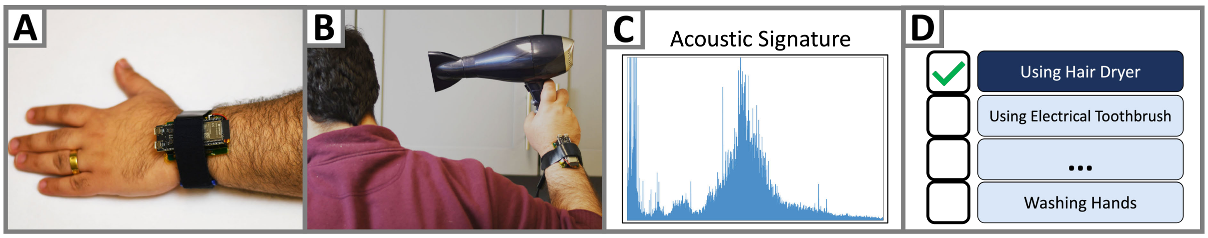

Fig. 1. System Overview of HandSAW: A) A customized PCB wearable system with 4 VPU sensors for surface acoustic wave collection and wireless transmission. B) User interacts with daily objects, a hairdryer as an example. C) Acoustic signal of daily objects has different signatures in the raw waveform. D) Real-time machine learning pipeline predicts the activities that the user is performing.

Enabling computing systems to detect the objects that people hold and interact with provides valuable contextual information that has the potential to support a wide variety of mobile applications. However, existing approaches either directly instrument users' hands, which can reduce tactile sensation, or are limited in the types of objects and interactions they can detect. This work introduces HandSAW, a wireless wrist-worn device incorporating a Surface Acoustic Wave (SAW) sensor with enhanced bandwidth and signal-to-noise ratio while rejecting through-air sounds. The device features a sealed mass-spring diaphragm positioned on top of the sound port of a MEMS microphone, enabling it to capture SAWs generated by objects and through touch interaction events. This custom-designed wearable platform, paired with a real-time ML pipeline, can distinguish 20 passive object events with >99% per-user accuracy and a 91.6% unseen-user accuracy, as validated through a 16-participant user study. For devices that do not emit SAWs, our active tags enable HandSAW to detect those objects and transmit encoded data using ultrasonic signals. Ultimately, HandSAW provides an easy-to-implement, robust, and cost-effective means for enabling user-object interaction and activity detection.

CCS Concepts: • **Human-centered computing → Interaction devices**; **Ubiquitous and mobile devices**; **Mobile devices**; • **Hardware → Tactile and hand-based interfaces**; **Sensors and actuators**; **Tactile and hand-based interfaces**; **Sensors and actuators**;

Authors' Contact Information: Kaylee Yaxuan Li , University of Michigan, Ann Arbor, Michigan, USA, yaxuanli@umich.edu; Yasha Iravantchi, University of Michigan, Ann Arbor, Michigan, USA, yiravan@umich.edu; Yichen Zhu, University of Michigan, Ann Arbor, Michigan, USA, ychenzhu@umich.edu; Hyunmin Park, University of Michigan, Ann Arbor, Michigan, USA, phyunmin@umich.edu; Alanson P. Sample, University of Michigan, Ann Arbor, Michigan, USA, apsample@umich.edu.

ACM 2474-9567/2025/3-ART10
https://doi.org/10.1145/3712276

Additional Key Words and Phrases: Surface Acoustic Waves, Object Interaction, Activity Recognition, Wearables, Sensing

ACM Reference Format:
Kaylee Yaxuan Li, Yasha Iravantchi, Yichen Zhu, Hyunmin Park, and Alanson P. Sample. 2025. HandSAW: Wearable Hand-based Event Recognition via On-Body Surface Acoustic Waves. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 1, Article 10 (March 2025), 29 pages. https://doi.org/10.1145/3712276

1 Introduction

Empowering intelligent systems to detect how people interact with objects can enable human activity recognition [15, 31, 50, 81, 87, 108, 123], automated health and wellness monitoring [25, 26], real-time guidance for tasks [24, 49], and immersive AR/VR experiences [46, 75]. To achieve this vision of hand-based event recognition, researchers have explored various sensing techniques including cameras [32, 67, 114], inertial measurement units (IMU) [14, 86, 87, 106], electromyography (EMG) [40, 70], pressure sensors [1, 17], and radio/acoustic transceivers [10, 13, 117]. Many designs have adopted glove-based configurations that directly instrument the hands for greater signal fidelity, but glove-based designs can be cumbersome for long-term wear. As a result, many recent approaches limit the instrumentation to rings or opt for wrist-worn strategies [48, 103]. These approaches often rely on techniques that permit signal transfer through the body, such as passive electromagnetic (EM) or active radio frequency (RF) sensing [53, 57, 108, 120]. However, these approaches are often limited in the types of activities or objects that can be detected, as they typically function only with objects that emit EM signals or are composed of compatible materials. While microphones have been explored for human activity recognition, including in wearable applications [10, 13], microphone-based methods may have difficulty segmenting the user's individual activities from simultaneous activities happening in the environment, and the "always-on" operation can invoke privacy concerns due to the nature of the information they collect, such as a bystander's speech. While recent work looks towards ultrasonic signals [118, 121, 129] to reduce privacy concerns, such approaches also reduce the scope of events these systems can recognize. Ideally, a wearable system should be able to identify a variety of objects, be easy to integrate into existing wearables, and be robust to environmental noise.

Given these constraints, we identified Surface Acoustic Waves (SAWs) as a valuable signal source for mobile and wearable applications. Many objects that have kinetic qualities—the moving gears in a drill, fluids running through a sink fixture—generate waves when they are activated: sound waves through the air, traditional 3D bulk vibrations through the material, and the aforementioned SAWs that travel along their 2D surface [71, 85]. Traditional vibrations and SAWs can travel from objects and into the body upon touch. This characteristic enables their respective sensing approaches to reject in-air environmental sounds and noises while allowing for the active selection of signals exclusively from objects the wearer touches. This paper presents HandSAW, a wearable platform that utilizes a new class of contact-based surface-acoustic-wave sensors originally designed for hearing aids and marketed by the manufacturer as a "Voice Pickup Unit" (VPU). It has characteristics significantly different from those of previous contact microphones. It consists of a hermetically sealed mass-spring diaphragm placed on top of the sound port of a standard MEMS microphone, which effectively eliminates sound transmitting through air and is sensitive to mechanical vibration signals from the contacted surface. This new device provides superior bandwidth (48kHz) and signal-to-noise ratio (SNR).

By capitalizing on the capabilities of the VPU, we propose to repurpose this device for on-body sensing for hand-based object and activity detection. When a user interacts with an object, the SAWs generated by the object or via contact with the object carry unique acoustic signatures that travel along the surface-to-air boundary of their body. By sampling these signals with the VPU, the wearable platform is able to classify the activity being performed. The HandSAW system is shown in Figure 1A. When a user touches an object, SAWs pass along the surface-to-air boundary of their body and into the wearable platform (Figure 1B). HandSAW's signal processing (Figure 1C) and machine learning pipeline are then able to classify user activities and events in real-time (Figure 1D). Given that not all objects emit acoustic waves when interacted with, an "active" communication is proposed as an alternative method for identifying "silent" objects. For these objects, low-cost ultrasonic transducers can be equipped to encode information using methods such as Amplitude Shift Keying (ASK). Bits 1 and 0 are programmatically generated, allowing HandSAW to decode them for data transfer when a user touches the object. Beyond augmenting objects with "active" ultrasonic transmitters, several common devices, such as smartphones, can also encode data through audible or inaudible sounds or vibrations. This dual approach can work in tandem with the passive signals that already exist or allow any object to be instrumented with an active tag to enable through-body SAW-based communication.

This work evaluates SAWs in wearable applications from four aspects. First, compared with two common wearable sensors (an accelerometer and piezoelectric disk), VPU maintained the highest SNR across a wide frequency range and offered the greatest consistency in reliably capturing signals across participants. These experiments demonstrated that the VPU can deliver strong recognition performance across a range of activities and across users from a theoretical operational perspective. Second, HandSAW was evaluated through a 16-participant user study of 20 classes of events, including tactile feedback, workshop, home, office, and self-care activities. We found >99% per-user accuracy, as well as an across-user accuracy of 91.6%, demonstrating strong performance even for a new user. Third, the performance for active data transfer encoded by ultrasonic signals with the ASK method was measured through a bit error rate evaluation, finding that HandSAW can sustain 3000bps with around 1% bit error. Finally, illustrative applications highlight the practical utility of HandSAW: wearable activities tracking daily living and encoded information flow using ultrasonic active data transfer.

This paper makes the following contributions:

(1) A wearable platform capable of capturing SAW-based object interaction events and receiving active data transfer;
(2) Experimental results showing HandSAW's passive sensing capabilities, which can recognize 20 different hand-object interaction events consistently across 16 participants;
(3) Assessment of HandSAW's active data transfer capabilities as the way filling gap for hand interactions that do not generate sound, which can sustain 3000bps with 1% bit error rate;
(4) Applications highlighting HandSAW's utility in wearable daily activity recognition and context-driven data transfer.

2 Related Works

In this section, we contextualize HandSAW by first providing a broad overview of sensing approaches related to the hands. We then focus on activity detection through wearable devices using acoustic methods.

2.1 Sensing Approaches for the Hands

Given the importance of users' hands in various forms of interaction with computers, objects, and each other, it is no wonder that a significant body of research within the mobile computing system literature is dedicated to enabling computers to better understand what our hands are doing. From hand pose [32, 37, 41, 42, 86, 106], discrete gestures [36, 47, 55, 67, 114], to hand-touch detection [43, 82, 109], hand-object interaction [28, 53, 57, 93, 101], human activity recognition [27, 43, 44, 95, 124, 125], hand-related sensing technologies generally fall into two main categories: environmental and wearable approaches. Environmental methods, primarily relying on optical technologies like overhead cameras and fiducial markers (e.g., Vicon system [98], CornerCameras [92]), have advanced with the growth of computer vision (CV) [6, 93, 94, 126]. While these systems offer the advantage of a one-time setup, they are limited by line-of-sight issues or become ineffective outside the instrumented environment. To overcome these limitations, the focus on human-computer interaction has shifted towards wearable technologies for human activity recognition (HAR). Sensing mechanisms such as acoustic [10, 13, 117], IMU [14, 52, 58, 69, 77, 96], and RF backscatter techniques [9, 57, 73, 83, 84, 100, 120] have been explored. Wearables, being in constant contact with the user, offer continuous tracking across various environments and interactions. Early wearable solutions like glove-based systems [102] provided high signal fidelity but were impractical for long-term use due to their intrusiveness and discomfort. This has led to the development of less obtrusive alternatives such as sensing rings, wristbands, and armbands [3, 21, 67, 68, 81, 97, 104], which capitalize on the hand's natural ability to conduct signals. For example, EM-Sense wristband [53] and Z-Ring [108], exploit the transmission of radio frequency (RF) and body signals through the hands to enable hand-based activity recognition. A detailed comparison with several typical sensing techniques is provided in Section 7.1 for a better contextualization of this research.

Our approach, HandSAW, follows a similar philosophy, using a wristband to capture Surface Acoustic Waves (SAWs) generated during hand-object interactions, thus offering seamless integration of hand sensing and wearable technology for activity recognition. Enhanced by the advanced design of the VPU sensor for capturing SAWs, HandSAW provides superior bandwidth and signal-to-noise ratio while rejecting through-air sounds. This work fundamentally expands the scope and breadth of interactions that can be detected while maintaining superior accuracy and robustness across diverse users and environments. It is important to note that the recognition of completely static hand gestures is outside the scope of this research.

2.2 Human Activity Recognition Using Acoustic Sensing

Most related to HandSAW in technique are acoustic approaches for sensing activity recognition. Most commonly, acoustic approaches use microphones, which capture in-air sounds, featurize them with Fast Fourier Transforms (FFTs) or Mel-Frequency Cepstral Coefficients (MFCCs), and use classical or deep learning approaches to classify acoustic events [10, 13, 105, 117]. Microphones may also be combined with active emitters over the air to perform Frequency Modulated Continuous Wave (FMCW) recognition of activities [60, 64, 112, 127]. However, always-on microphones create privacy concerns, which may dissuade users away from continuous use, especially in wearable settings [35**?**]. Some FMCW systems operate entirely with ultrasonic emitter and receiver pairs and do not capture audible frequencies, such as speech content [61, 72, 76, 89, 118, 122, 129]; this also means they cannot passively capture signals emitted from objects the user interacts with. To passively capture signals from objects without capturing audible or speech content[88, 115, 116, 121, 128]. SAWSense [38] utilizes the latest generation of VPU microphones from Sonion Inc. to capture SAWs from physical objects' operating surfaces. However, SAWSense does not have the capability to distinguish how a user activated the devices or which object is currently being touched and interacted with. In contrast, HandSAW captures signals through user touch, which enables the system to recognize exactly when and even how a user is interacting with objects. For example, the system can identify actions such as using a Dremel tool at a low speed.

3 On-Body Sensor Comparisons

This section introduces the background of Surface Acoustic Waves (SAWs) and the design details of Voice Pickup Units (VPU). Subsequently, experiments compared the VPU with two common wearable sensors—an accelerometer and a piezoelectric disk. The VPU demonstrated superior SNR across a wide frequency range and reliably captured signals across participants, establishing a strong foundation for future wearable device development.

3.1 Primer on Surface Acoustic Waves and Voice Pickup Units

Many objects with kinetic qualities, such as the motor in a hand drill or the levers in a doorknob, create waves upon their activation through the air (sound), the 3-D bulk of the material (bulk vibrations), and along their 2-D surfaces (surface acoustic waves) [71, 85]. When these waves emanate, traditional vibrations and SAWs can travel through the hands and into the body upon touching these objects. This quality, where vibrations and SAWs only travel on the body on touch, allows their respective sensing approaches to reject in-air environmental sounds and noises. This quality also enables active selection in only capturing signals from objects the wearer touches. However, given that SAWs propagate in 2 dimensions, the amplitude decay is reduced when compared to 3-D bulk vibrations ($1/\sqrt{r}$ vs. $1/r$) [45, 91]. This lower decay rate suggests that a larger fraction of the original signal will reach the hands, thus offering a greater ability to sense more subtle events through SAWs.

Based on these beneficial characteristics of SAWs, we utilize Sonion Voice Pickup Units (VPU) to capture these signals. The VPU was initially designed as a compact ($3.5 \times 2.65 \times 1.50$ mm) bone conduction microphone for hearing aids optimized to capture a wearer's voice without capturing environmental noise through the air [99]. Functionally, the VPU contains a mass-spring diaphragm mounted on top of the sound port of a standard INVN (ICS-40619) MEMS microphone, which are both hermetically sealed inside an enclosure to prevent outside sounds from being captured by the microphone. The design is impedance-matched for sensing direct contact vibrations rather than free-air acoustic vibrations. By adding mass to the diaphragm, its resonant frequency can be controlled, and the Quality factor or "Q" of the resonator can be tuned to improve the VPU's sensitivity and maintain the wide bandwidth of the MEMS microphone. As a result, the VPU from Sonion is able to detect SAWs through direct contact while robustly rejecting in-air noise, which we have repurposed as a high bandwidth and high signal-to-noise ratio general-purpose SAW sensor.

The use of on-body SAWs has been explored primarily in the medical literature [56]. Earlier studies have shown that the elastic properties of cell membranes allow SAW signals to propagate efficiently through human tissue [74]. These properties have led bioengineers to innovate in areas like measuring elasticity in soft materials or human skin [39], detecting melanoma through laser-generated SAWs [12], isolating stem cells, creating cell spheroids, and manipulating cells within scaffolds [39]. However, while these works strongly support that SAWs can travel well when actively generated on the body, we use this section to evaluate how well passive signals from objects maintain these properties and travel along the body to be captured by Voice Pick Up Unit (VPU) sensor. Additionally, as a point of comparison, we include two common sensors used in on-body sensing applications: an accelerometer and a piezodisk transducer.

3.2 Experimental Setup & Sensor Selection

Although prior work [38] compared VPU and other common sensors, including accelerometers, microphones, and geophones, they were evaluated by placing the signal source and sensor on the same surface. This differs from wearable situations where the sensor is in contact with the user's body (such as on the wrist), and the signal has to "jump" from the object to the body. Furthermore, while signals may have similar signatures on materials within the same categories (i.e., similar glass and metals), achieving consistent performance across different users can be challenging. Acoustic signals, in particular, can vary significantly from one person to another [37]. Several different sensors are considered for comparison. In reference to traditional surface transducers, the two most commonly used types are geophones [30] and piezoelectric transducers [7, 33]. While widely used, geophones suffer from extremely limited bandwidth (usually less than 250 Hz), are bulky, and are limited to vertical oration for gravity due to the nature of their design. These factors make them fundamentally unsuitable for wearable and mobile applications. Traditional contact microphones, such as the piezoelectric transducer (e.g. Knowles BU-21771), are piezoelectric sensing devices in surface-mount packages, which operate on the same principles as the piezoelectric disk. Considering Knowles BU-21771 costs $61 USD, a standard piezoelectric disk is selected for comparison. Apart from privacy concerns, microphones capture environmental noise when used in an "always-on" fashion, which is not considered for wearable use. Moreover, these factors warrant additional evaluation of VPU at various locations on the body with diverse participants.

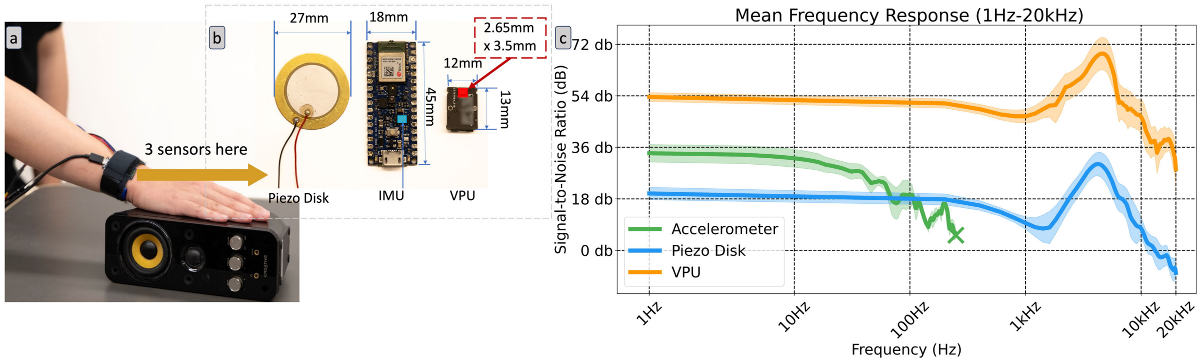

Fig. 2. (a): Frequency response test setup includes a speaker connected to a function generator. (b): Participants placed their palms on the speaker with three sensors—IMU, piezoelectric disk, and VPU—on the wrist for synchronized data collection. Note: VPU dimensions are 2.65mm × 3.5mm, PCB size 12mm × 13mm. (c): Depicts mean frequency response curves for the three sensors from 1Hz to 20kHz across four users, with an 'X' marking the accelerometer's Nyquist limit.

As a point of comparison, we selected LSM9DS1 [62] inside an Arduino Nano Sense 33, and add the 7BB-27-4L0 piezoelectric disk [19], as shown in Figure 2(a). The output data rate (ODR) of the LSM9DS1 was set to 500Hz, representative of high-end accelerometers in consumer-off-the-shelf (COTS) devices. Although IMUs with greater bandwidth and higher ODRs exist, such as ones in the kilohertz range [18, 34], there are a few factors we considered when selecting the LSM9DS1 and setting its ODR. First and foremost, the LSM9DS1 ($6) is priced roughly on the same scale as the VPU ($2) and piezoelectric disk ($0.85), compared to the specialized high-frequency low-noise ADXL1001 ($75). Second, while comparatively cheap COTS IMUs can be set to significantly higher ODRs, their sensor bandwidth is often limited due to underlying mechanical limitations and Digital Low Pass Filters (DLPF) that ensure the output of high SNR data [34]; in this case, the bandwidth is not necessarily half the ODR. For example, while the MPU-9250 (a COTS-grade IMU) can be set to an ODR of 4kHz, a 1.1kHz bandwidth can only be achieved by disabling the DLPF. Thus, for IMUs, often a higher ODR does not imply increased bandwidth and comes at the cost of additional noise, thus compromising the data quality even when only capturing lower frequency signals [18]. The LSM9DS1 has a sensor bandwidth of 400Hz with its DLPF enabled, ensuring low noise performance when its ODR is set to 500Hz. In this study, a piezoelectric disk was interfaced with a Behringer UM2 [4], configured at a 48kHz sampling rate with a 200x gain preamplifier and 1 Megohm impedance input. The VPU has a digital Pulse Density Modulation (PDM) interface configured with an output audio signal bandwidth of 0-24kHz. The experimental setup is situated within an acoustically sealed chamber (2m × 2m), minimizing both noise and vibrations from the outside. The experiment was conducted on four participants (two females and two males), with an average age of 24.0 (SD=1.9). The mean wrist diameter was 16.4cm (SD=0.8cm), and the average arm length was 68.8cm (SD=0.8cm). We note that these evaluations were performed in accordance with our institution's IRB.

3.3 Frequency Response Evaluation

We first conducted the following tests, which examined frequency response of different sensors when attached on the body, to understand sensors' capability to capture various frequency acoustic signals. The frequency response of the VPU, accelerometer, and piezoelectric disk was assessed with the sensors concurrently mounted on the dominant wrist of participants, as shown in Figure 2. The participant's hand was placed on a GigaWorks T20 speaker housing, connected to a function generator performing a 1Hz to 20kHz linear sweep. Contact was limited to the hand and speaker, avoiding any other body or table contact. The sweep was simultaneously recorded by

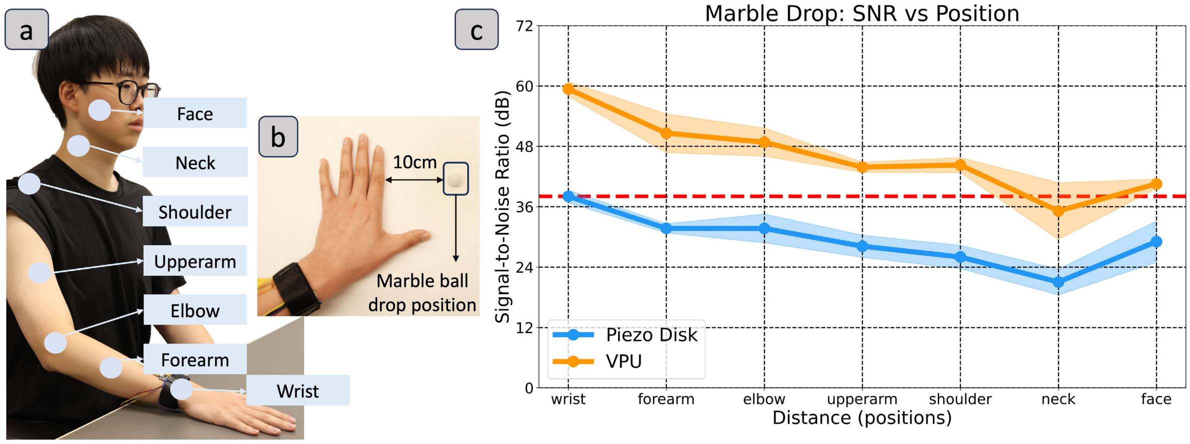

Fig. 3. (a): Experiment setup involved sensor testing on seven body sites: wrist, forearm, elbow, upper arm, shoulder, neck, and face. (b): A marble was dropped onto a table 10 cm from the user's hand, with sensors attached for data capture. (c): The study presents mean SNR curves for four users, comparing the VPU and piezo disk across the seven body locations, while the IMU was excluded due to its high noise levels.

all sensors. Signal-to-Noise ratio calculations across the sweep compared the captured signal at each frequency against the respective sensor's noise floor. Figure 2 illustrates each sensor's mean frequency response curve and standard error across participants.

Figure 2 reveals that the accelerometer effectively surpasses the noise floor but is restricted to frequencies below 250Hz due to its sampling rate. In contrast, the piezoelectric disk outperforms the accelerometer at higher frequencies, remaining above the noise floor up to nearly 10kHz. The VPU, however, consistently registers signals well above the noise floor across the spectrum, with its SNR dropping below 36dB only at 16kHz. Notably, the SNR of the other sensors never surpasses 36dB. All sensors exhibit low standard error across participants, with the VPU displaying the lowest average, indicating its superior consistency and potential for strong performance across different users.

3.4 On-Body Locations Evaluation

Following the demonstration of the VPU's capacity to capture a broad frequency range at the wrist with high fidelity, this study now examines signal propagation along the body, informing potential sensor placements for on-body sensing. In this study, impulses were generated by dropping a 5.5-gram, 1.5-cm diameter glass marble from 6 cm height, 10 cm away from the hand positioned between the index finger and thumb (Figure 3(b)). This method was replicated at seven body locations (wrist, forearm, elbow, upper arm, shoulder, neck, and face), as illustrated in Figure 3(a). Participants maintained hand-only contact with the table, mirroring the earlier evaluation. The marble drop experiment, selected for its controlled reproducibility and real-world applicability, is ideal for studying impulse signals. Such signals, characterized by short, high-energy bursts, have a wide frequency range according to the Fourier Transform principle. This makes a marble drop, as an almost ideal impulse signal, useful for analyzing broad-spectrum frequency responses.

Figure 4 shows the IMU's inability to clearly capture the marble drop at the wrist, a pattern consistent across all participants and seven locations, due to its high noise floor. In contrast, both the VPU and the piezoelectric disk distinctly registered the marble drop. Nonetheless, the piezoelectric disk's high noise floor limited its detection of

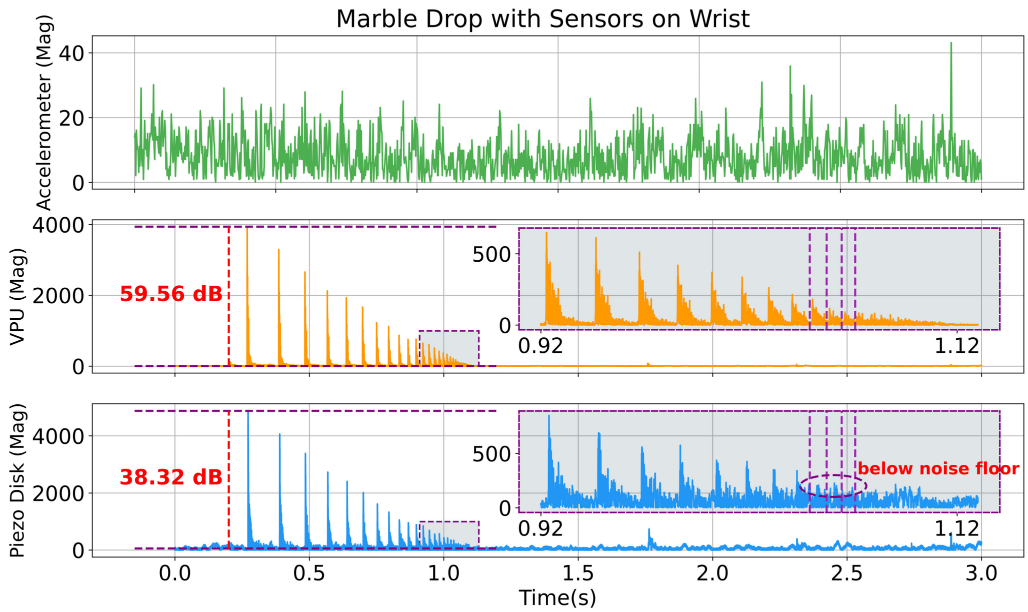

Fig. 4. Raw ADC data of three sensors in the marble drop experiment. The VPU effectively distinguishes between signal and noise in the final diminishing bounces, whereas the piezo disk mainly registers noise over the actual signal towards the end, indicating superior performance of the VPU in capturing faint signals. The SNR is also larger for VPU than Piezo.

finer bounces, unlike the VPU. The VPU's superior sensitivity enabled high-fidelity capture of the marble drop, even on the face, as evidenced by the mean SNR curves (Figure 3). Except for the neck, the VPU consistently outperformed the piezoelectric disk, which showed peak performance at the wrist. At the neck, both sensors' performance declined, likely due to the inadvertent capture of respiratory sounds, increasing the noise floor.

3.5 Environmental Evaluation

Previous evaluations established the VPU's superiority over accelerometers and piezoelectric disks in frequency response and body placement under optimal conditions. Extending these tests to more realistic scenarios, we introduced background music during the marble-drop procedure, as depicted in Figure 5. The accelerometer, which could not detect the marble drop in any test, was excluded from this evaluation. The piezoelectric disk exhibited a notable noise floor increase when music played, indicating its sensitivity to airborne sounds. Consequently, we conducted a modified marble drop test with the participant's wrist elevated, holding both sensors in the air. In this setup, the piezoelectric disk registered the marble drop through the air, albeit with reduced amplitude, while the VPU did not. The majority of piezoelectric disks used in object interaction and surface detection research are designed to operate in the lower range of human hearing. Generally speaking, a piezo disk consists of a flat surface made of piezoelectric material that generates an electrical charge in response to mechanical stress. This allows the disc to detect mechanical waves and convert them into corresponding electrical signals. However, for the piezo disk, various sources can induce mechanical stress and thereby cause it to capture unintended signals. For example, for a given piezo disk, airborne sound waves can induce vibrations in the disk's structure, allowing

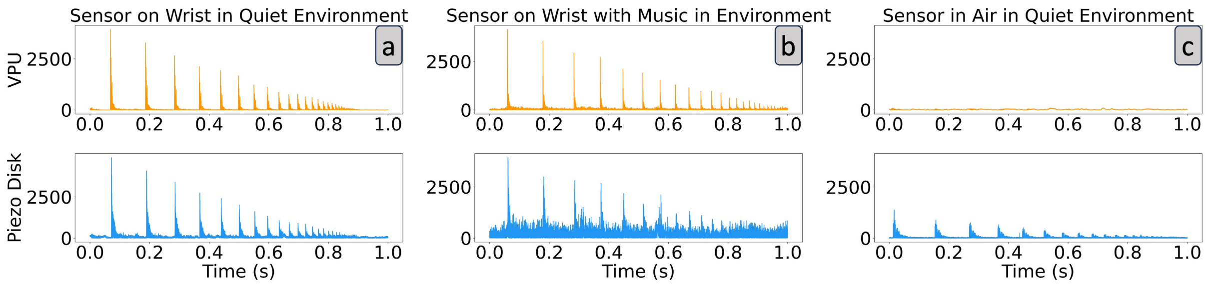

Fig. 5. Signal comparison between VPU and piezo disk under three conditions: (a) marble drop on a table with sensors on the wrist in a quiet environment, (b) same setup with background music, and (c) sensors held in the air in silence. These scenarios indicate the VPU's resistance to environmental noise, contrasting with the piezo disk's sensitivity, essentially functioning as a microphone.

it to capture sound from the air. In contrast, VPU is designed to be impedance-matched for direct contact with a surface. This means they are optimized for a different acoustic velocity-to-pressure wave ratio than a piezoelectric disk. This is achieved primarily by a mass-spring diaphragm placed over the sound port of a MEMS microphone and hermetically sealed in a surface-mount package, preventing acoustic energy from entering the SAW sensor's package. As a result, this design is highly sensitive to vibrations when in direct contact with an object. These evaluations collectively demonstrate the VPU's robust signal fidelity across various frequencies and different body locations, along with its ability to effectively reject environmental noise. These results establish the VPU as a method of robustly detecting physical contact-based signals for further wearable system design.

4 System Implementation

Building off of the results from Section 3, the VPU has been selected given its excellent through-air sound rejection, excellent signal-to-noise ratio, wide bandwidth and ability to detect Surface Acoustic Waves as they travel across the body. This section provides an overview of the hardware and software design of HandSAW. In addition to implementation details, this section details various design considerations for the wearable capture of SAWs, a featurization schema to support a wide range of activities, and an encoding scheme for active data transmission.

4.1 Hardware Design

HandSAW utilizes Sonion Voice PickUp (VPU) sensors initially designed to in-ear pick up the wearer's voice through the body without capturing environmental noise, such as in a crowded location [11, 90]. They utilize an industry-standard Pulse Density Modulation (PDM) scheme to transfer data digitally and can be paired to share the same clock. By selecting the ESP32-S3 microcontroller, which has an integrated I2S (Inter-IC Sound) hardware peripheral [20], the system can efficiently decode this stereo PDM stream. As seen in Figure 7, HandSAW utilizes one peripheral with four data lines, each decoding a mono-microphone, resulting in 4 sensing channels at a 16-bit 48kHz sampling rate. The customized PCB board has been designed featuring two VPUs soldered directly onto the board, with the option to connect two additional VPUs via connectors as illustrated with Figure 6(a). These sensors are distributed around the wrist, as seen in Figure 6(c,d), and are later used to determine whether the position or number of sensors affects system performance in Section 5. The PCB board includes a hollowed-out section for Velcro integration, allowing the wristband to conform to the geometry of the wrist. This design ensures that the VPUs maintain good contact with users of varying arm sizes and can be worn comfortably over long periods. The ESP32-S3's WiFi module transmits the datastream via socket to a laptop for data collection.

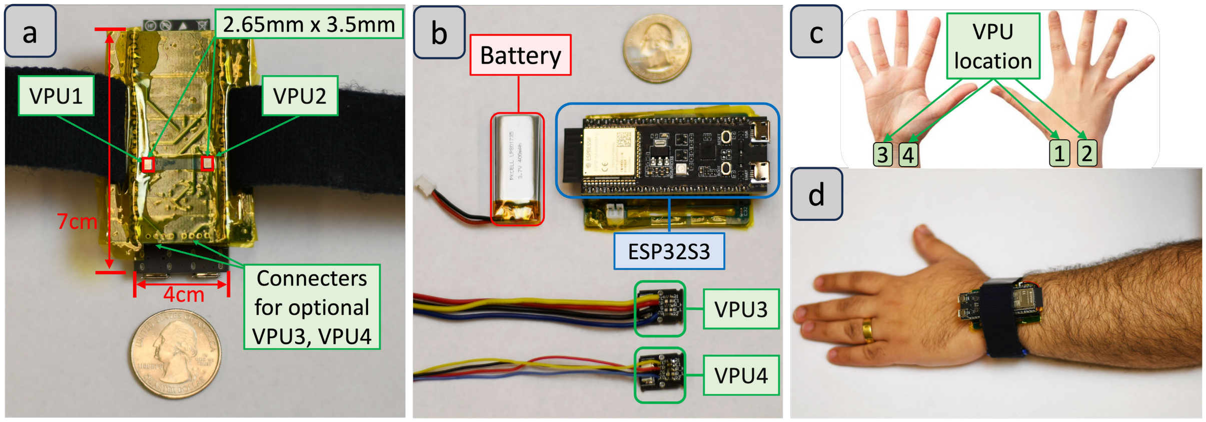

Fig. 6. Overall hardware device setup for HandSAW. (a): Customized PCB board's backside featuring VPUs and a hollowed-out section for Velcro integration. (b): Frontside of PCB with ESP32 plugged in. The VPU3 and VPU4 PCB boards. 3.7V batteries placed between ESP32 and PCB. (c): The positions of four VPUs on the wrist. (d): User wearing the system.

A 3.7V 400mAh lithium-ion battery (36mm × 17mm × 7.8mm, 8.2g) provides enough power for approximately 5 hours of continuous operation, allowing for fully wireless functionality, as illustrated in Figure 6(b). Detailed power analysis was conducted under different modes of ESP32S3. In deep sleep mode with 8.14×10^{-6}W, HandSAW conserves energy by powering off the CPU, majority of RAM, and digital peripherals synchronized with APB_CLK. The PDM to PCM mode with 0.15246W turns off WiFi, keeping only VPU data reception active. In WiFi Mode with 0.30855W, the audio peripheral I2S is disabled, enabling WiFi to transmit data equivalent to four VPUs' output (96000 × 4 data points per second). For testing HandSAW running mode with 0.4224W, the device operates normally without power-saving configurations, utilizing both the audio peripheral and WiFi module transferring 20488 data every 106ms.

4.2 Software Design

Upon receiving data via socket stream as mentioned in Figure 7, the system employs a 1-second sliding window with a 0.1-second step size for enhanced responsiveness and to prevent event segmentation across multiple windows. A Short-Time Fourier Transform (STFT) is applied to eliminate silence within the window. The remaining audio is then processed using a custom Mel-Frequency Cepstral Coefficients (cMFCC) method, facilitating real-time predictions by the machine learning model, as depicted in Figure 7. While the Fast Fourier Transform (FFT) is a feasible method for audio feature extraction, its comprehensive feature set may impede real-time system performance due to extended processing times. Conversely, Mel-Frequency Cepstral Coefficients (MFCCs) provide a faster choice. It offers advantages in reducing signal dimensionality and improving computational efficiency compared to FFT without significant loss of information. However, traditional MFCCs, primarily designed for human auditory perception, may not represent non-speech acoustic signals optimally. Additionally, standard 20 MFCC bins distribution is inadequate for Surface Acoustic Waves (SAWs) [38]. To overcome this, a customized MFCC approach was developed using the librosa library [59]. An initial study with two participants across 13 classes showed that 128 mel filter banks distributed from 0Hz to 24kHz yielded the most balanced performance across various events captured. For 1 second of data, 128 feature arrays of each channel are concatenated. Therefore, the total number of feature lengths of 1-second input data depends on the channel count.

Feature extraction is followed by event classification using a machine learning model on a laptop. Three different approaches are employed: (1). A linear Support Vector Machine (SVM) using SciKit-Learn's default

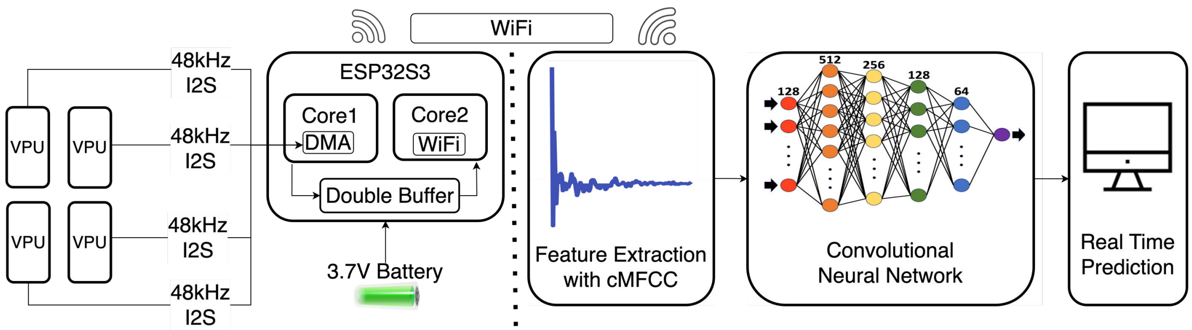

Fig. 7. The block diagram includes: (a) ESP32S3 connected to four VPUs with dual-core processing, and (b) ESP32S3's wireless WiFi communication to a computer for server-side cMFCC feature extraction and machine learning-based real-time predictions.

parameters serves as a computationally efficient baseline, robust against overfitting and indicative of potential on-device deployment in the future. (2). A Random Forest, also in SciKit-Learn, with 250 estimators and default parameters, excels in handling complex feature relationships in datasets, beneficial for distinguishing similar classes with overlapping features (e.g., a drill operating at a low vs. high speed). (3). A TensorFlow-based deep neural network, with four fully connected layers, is designed for large, multi-user datasets. This neural network architecture accelerates training integrated with batch normalization, introduces non-linearity via ReLU activation for enhanced error learning, and reduces overfitting using a 0.5 dropout rate. It concludes with a softmax layer, computing probabilities for all events enabling accurate activity classification. Ultimately, this system facilitates real-time classification of user activities and events. The average latency for ten iterations of each step in the pipeline was calculated: WiFi latency is approximately 30ms, and ML-model inference times are 34ms for the deep learning model, 1.34ms for the SVM model, and 1.49ms for the RF model. The model's size, at about 400KB for the model .pkl file and 200KB for the parameters .json file, is sufficiently small to allow operation directly on mobile computing devices, eliminating the dependence on server-based processing. Further enhancing HandSAW, RF and neural network were deployed on the ESP32S3 board. The implementation first used the emlearn library [79] to convert the model, originally trained in Python, into pure C code. This C-based model was then deployed onto the ESP32S3 module to produce predicted labels. The predicted labels can either be saved as a logging file on the board for later download to a computer or sent immediately to a server via WiFi.

5 System Evaluation

In this section, we investigate HandSAW's proficiency in recognizing passive events across three environments (office, workshop, kitchen) and two applications (tactile feedback, personal care), using a sample of 16 participants. Our methodology involves three distinct evaluations: (1). Per-User performance, analyzing the system's accuracy when trained on data from individual users. (2). Across-user performance estimates the system's performance on a new user, where the system is trained on all but one participant and tested using the remaining participant's data. (3). The effect of the position and number of sensors on performance. All evaluations were carried out in compliance with our institution's Institutional Review Board (IRB).

5.1 Class Environments and Scenarios

Our evaluation of HandSAW's versatility encompassed three environments (office, workshop, kitchen), chosen based on Section 2. In the office, activities included keyboard typing, writing with a pen, turning a door knob,

Fig. 8. 20 classes of hand-based events with corresponding frequency distribution are shown in FFT graphs. Please note that the upper limit on the y-axis has been adjusted for visualization purposes, as the shape of the curve is more informative than the actual amplitude of the values.

and using a compressed-air spray duster. Workshop tasks involved common tools such as a hammer, a cordless drill, and a Dremel at two speed settings (low and high). In the kitchen environment, we selected two appliances, a microwave and blender, and two interactions with fixtures, opening a refrigerator door and pulling a paper towel from a dispenser. Additionally, we identified two common scenarios for hand-based event recognition: tactile feedback and personal care activities. For tactile feedback, we selected three gestures that benefit most from uninstrumented hands and the resulting tactile sensation when performed: tapping on a surface with a finger, clapping both hands and swiping a finger on an opposing arm. For personal care activities, we selected important hygienic and wellness activities such as using a toothbrush, washing hands, using a hairdryer, and using an electric massager. The frequency signatures of these diverse activities are illustrated in Figure 8.

5.2 User Study Procedure

We recruited 16 participants (8 female, 8 male; mean age = 22.7, SD = 1.8) to wear HandSAW on their dominant hand and perform the 20 activities detailed above 50 times in a randomized order to ensure diverse data collection and avoid repetitive patterns or temporally similar samples. Participants performed these activities naturally, with initial guidance, but participants were encouraged to adapt and perform them according to their natural tendencies, leading to slight variations in technique. The study was conducted in a dynamic workshop environment with background activities like conversations, movements, building HVAC air conditioning, and a 3D printer sometimes operating, reflecting real-world conditions with no adjustments to reduce these potential noise sources. The

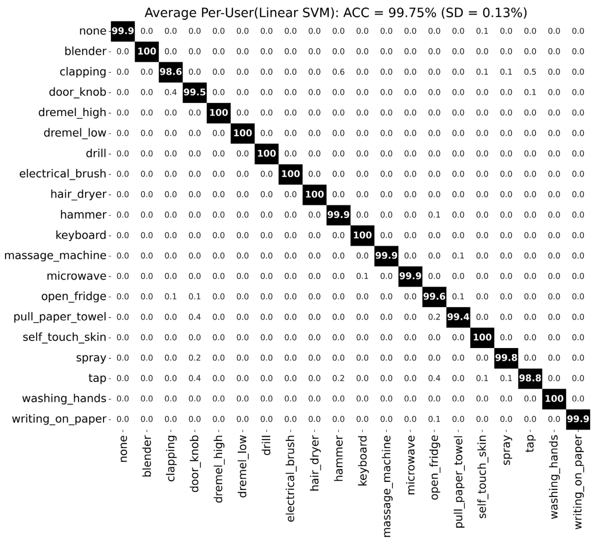

Fig. 9. Per-User confusion matrix that shows the mean classification accuracy of 16 users for the 20 different classes using the Linear Support Vector Classifier.

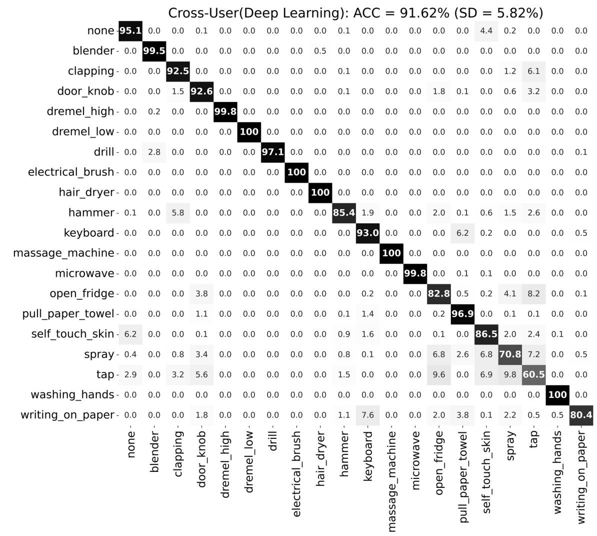

Fig. 10. Cross-User confusion matrix shows the deep learning model's mean accuracy using leave-one-participant-out (LOPO) on left-out unseen users.

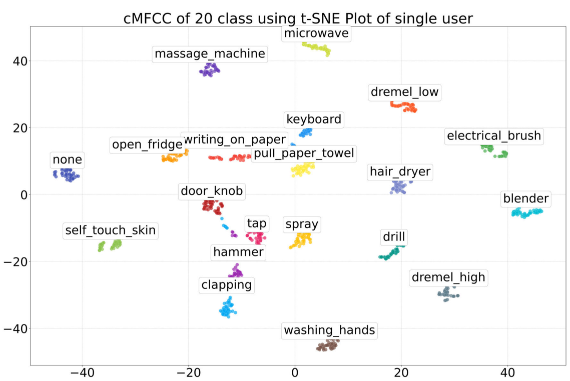

Fig. 11. The t-SNE plot visualization for features of 20 classes of one user.

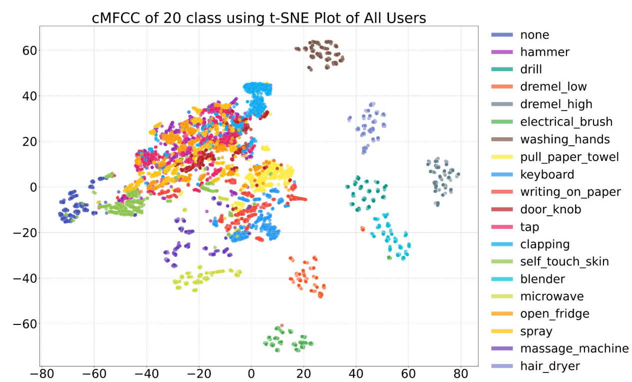

Fig. 12. The t-SNE plot visualization for features of 20 classes across all 16 users.

system's fully wireless, battery-powered nature allowed participants to engage in tasks comfortably, either sitting or standing. Acknowledging the study's extent (1000 actions per participant), breaks were permitted when taking off the device. Each participant's involvement spanned approximately two hours, with the entire dataset collected over two weeks.

Variations in our study arose from tool performance changes due to battery depletion of a cordless drill and electric brushes, as well as diminishing air pressure in four different spray bottles used by the 16 participants. Additionally, participant-specific approaches introduced variabilities, such as varying force levels in keyboard typing and different techniques in tapping actions (using fingernails or finger pads).

Table 1. The classification performance for motion recognition using 1,2,4 VPU channels for training across various machine learning classification approaches for per-user and cross-user evaluation. For cross-users, the LOPO experiment was conducted.

Channels \ Model	Support Vector Machine	Random Forest	Deep Learning
1 channel per-user	**99.24% (SD=0.42%)**	98.63% (SD=0.82%)	95.06% (SD=2.41%)
2 channels per-user	**99.62% (SD=0.08%)**	99.45% (SD=0.13%)	97.43% (SD=0.36%)
4 channels per-user	99.75% (SD=0.14%)	**99.76% (SD=0.18%)**	98.82% (SD=0.94%)
1 channel cross-users	84.89% (SD=1.6%)	83.06% (SD=1.75%)	**86.95% (SD=0.92%)**
2 channels cross-users	88.31% (SD=0.79%)	85.20% (SD=1.05%)	**91.83% (SD=0.42%)**
4 channels cross-users	89.75% (SD=5.62%)	86.49% (SD=7.33%)	**91.62% (SD=5.82%)**

5.3 Per-User Classification Performance

For HandSAW's per-user accuracy, we performed a 10-round cross-validation for each of the 16 users. In each round, we divided the user's data into 10 rounds, trained on nine, and tested on one without shuffling. The results were averaged for all combinations. We then calculated and reported the mean per-user accuracies across participants for the linear SVM, Random Forest, and Deep Learning model as described in Section 4.2.

Linear SVM achieved a mean per-user accuracy of 99.75% (SD = 0.14%), closely followed by the Random Forest with 99.76% (SD = 0.18%). The Deep Learning model, however, showed a lower accuracy of 98.82% (SD = 1.12%), likely due to the limited dataset size for individual users. Detailed results are presented in Table 1, and the corresponding mean confusion matrix is depicted in Figure 9. To further contextualize our results, we conducted a t-distributed Stochastic Neighbor Embedding (t-SNE) analysis using data from a single participant, as shown in Figure 11. t-SNE, a dimensionality reduction technique for visualizing high-dimensional data, helped reveal clusters or patterns. The analysis indicated minimal class overlap, aligning with the strong performance of all three models. Interestingly, classes with inherent variabilities, like writing force or toothbrush battery level, had their variety represented in the plot with sub-clusters within their clusters.

5.4 Cross-User Classification Performance

To evaluate how well HandSAW can recognize events independent of the wearer, we performed a leave-one-participant-out (LOPO) evaluation, where we trained each machine learning model on data from all but one user and tested it on the remaining user. This approach revealed a mean accuracy of 89.8% (SD=5.62%) for the linear SVM, 86.49% (SD=7.33%) for the Random Forest, and 91.45% (SD=6.65%) for the Deep Learning model, averaged across 16 participants. The confusion matrix for the Deep Learning model, detailed in Figure 10, shows that out of 20 classes, 8 had accuracies over 99%, while only 6 fell below 92%. In some classes, like hammer use or refrigerator opening, lower performance is attributed to user-specific significant variations in action execution. Notably, the spray duster's performance varied due to its four replacements and inconsistent air pressure. Additionally, a t-SNE analysis of all 16 users' data, illustrated in Figure 12, revealed distinct clusters for classes with high accuracy, underscoring consistent classification despite diverse user inputs. Conversely, classes with lower accuracies in the confusion matrix and ones we observed having significant variation across users represented that variation in the t-SNE plot; they did not form distinct clusters or overlapped with classes with other clusters, which matches the performance seen in the confusion matrix.

Finally, we evaluated the three models using varying sizes of training data. First, we selected a specific number of users, n, for the training set and then tested the remaining unseen users individually, computing the average classification accuracy from the results. Users were randomly chosen for training for each increment. This process

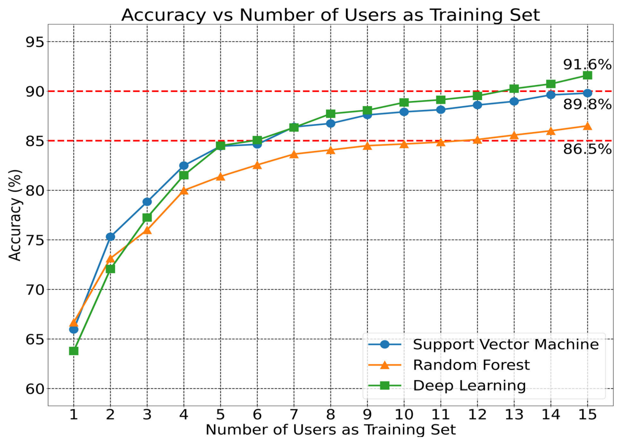

Fig. 13. The accuracy curves reflect performance on unseen user based on the number of users in training.

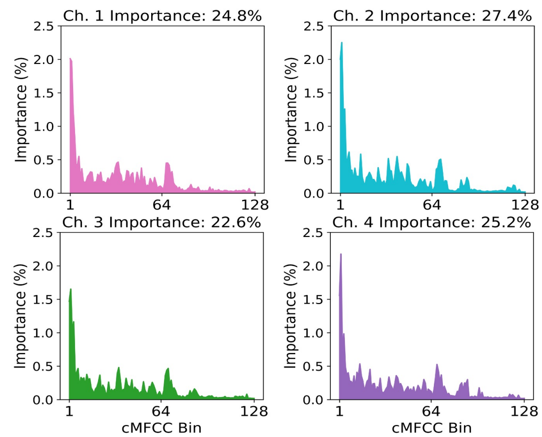

Fig. 14. Gini purity analysis assesses the relative importance of each channel.

was repeated for each model to determine the final mean accuracy over the number of users in the training set, represented as a curve. As shown in Figure 13, the consistent increase without plateauing suggests that enlarging the training dataset, potentially with augmentation, might narrow the gap between per-user and across-user accuracies. Achieving approximately 85% accuracy with only about 8 users in the dataset suggests near saturation of the training set. This indicates strong model generalizability and effectiveness with new, unseen users, confirming minimal variance in frequency response.

5.5 **Channel Importance For Classification Performance**

Finally, we investigated whether all four VPUs are crucial for HandSAW's observed high accuracy in activity recognition. We repeated the per-user and across-user evaluations with the three models, using configurations of one, two, and four channels/VPUs. As shown in Table 1, the results indicate that although performance generally increases with more VPUs, a single VPU configuration still demonstrated strong recognition capabilities. Notably, in per-user performance using a linear SVM, there was virtually no difference, and across-user performance reached 86.95% accuracy.

We further explored each channel's contribution through a Gini impurity analysis of the importance of cMFCC features across the four channels. Figure 14 shows that the channels had similar importance, with feature importances of 24.8%, 27.4%, 22.6%, and 25.2%, respectively. Additionally, the uniform importance distribution across all cMFCC bins suggests that each channel captures similar spectral information. This supports the finding that performance is comparable across channels, and even a single VPU can yield significant recognition performance as listed in Table 1. It also implies that VPU placement on the wrist is unlikely to impact system performance significantly, offering designers flexibility in sensor integration. However, for subsequent application development, we continue to use the four-channel version, as having multiple channels significantly increases the likelihood that at least one microphone will contact the skin. Utilizing more than one VPU enhances the robustness of the contact and provides reliable and stable output.

5.6 Recognition of "Silent" Activities via Active Data Transfer

Thus far, HandSAW has primarily focused on detecting events by capturing SAWs that emanate from objects during their operation (such as a drill running) or are generated through touch actions (such as a user tapping

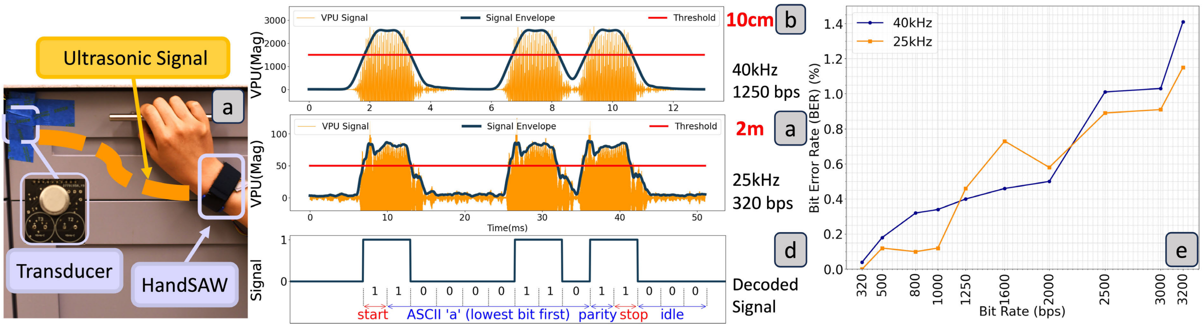

Fig. 15. HandSAW's active sensing system. (a): The user wearing HandSAW touches the drawer equipped with a transducer. The raw signal received by HandSAW is displayed: (b). The signal was generated by a 40kHz transducer transmitting at 1250bps and positioned 10cm away from the hand. c): Signal was generated by a 25kHz transducer transmitting at 320bps and positioned 2m away from the hand. (d): decoded message using UART protocol from the received signal. (e): Bit error rates performance with UART protocol at various bit rates, with 40kHz and 25kHz transducers using ASK.

on glass). However, one limitation of this approach is detecting events that do not generate signals during the event, such as softly placing a hand on an object or surface. To address this gap, we instrument these objects with "active" transducer tags that transmit the object's identity for "silent" objects.

Previous studies have confirmed the human body's effectiveness as a medium for signal transmission [2, 9, 29, 107, 111, 119, 119]. The human body effectively functions as an antenna across a wide frequency range when it is exposed to electromagnetic fields [23, 78]. The human body could resonate with a frequency that spans from 40 Hz to 400 MHz, attributable to the body's properties as a dielectric with intricate structure [22]. Previous research has extensively explored applications that leverage this characteristic of the human body. For example, IDSense and EMSense utilize RFID or electromagnetic technologies to detect interactions between humans and objects [53, 57]. Meanwhile, other researchers employed the human body as a receiving antenna for electromagnetic noise to identify touch locations on a wall in the home [16]. However, rather than passively recognizing signals, HandSAW actively controls and modulates signals to transmit data to wearable devices through physical human touch.

5.6.1 Implementation. In the previous sections, HandSAW's VPUs were configured to standard PDM clock and decimation ratios, resulting in a 48kHz sampling rate (i.e., a bandwidth of 0Hz-24kHz). While this configuration effectively captures passive SAWs, a considerable portion of the frequency range overlaps with human hearing, limiting its effectiveness for carrier frequencies, as tag operation would generate audible sounds, potentially disturbing users. Thus, the active tag is ideal for operating outside human hearing while utilizing cheap and readily available transducers. Therefore, we overclocked the VPU to extend the sampling rate to 96kHz (0Hz-48kHz bandwidth), which would also provide the added benefit of faster data rates. We identified the VPU's clock limit at 8 MHz, beyond which it no longer provided a valid PDM output. Considering the peripheral constraints of the ESP32S3, we selected a sampling rate of 96 kHz and a decimation ratio of 64, yielding a final clock of 6.14 MHz—within the operational limit but exceeding the VPU's specified 3.072 MHz limit. An expected and observed tradeoff is an increase in the noise floor.

25kHz or 40kHz transducers as active tags were connected to a function generator, which generates the appropriate carrier frequency at $30V_{pp}$(Figure 15(a)). The function generator receives a modulation signal via an Arduino Uno, which modulates the data with Amplitude Shift Keying (ASK). In this scheme, a '0' and '1' bit are represented by 0V and 30V amplitude, respectively. As illustrated in Figure 15(b, c), a digital bandpass filter is applied to first extract the active signal. Then, a Hilbert transform (to extract the envelope), and a Savitzky-Golay (Savgol) filter (for smoothing) were applied. To decode the values, a threshold is set to define a “0” or “1” bit. Data was transmitted using the UART protocol as shown in Figure 15(d).

5.6.2 Evaluation. To evaluate performance, data rates ranging from 300 bps to 3200 bps were implemented for each transducer. A random ASCII pattern was transmitted to measure the Bit Error Rate (BER). A user wearing HandSAW positioned their hand 10 cm from the transducer, with 300 packets transmitted in each scenario. The average BER is calculated and presented in Figure 15(e). Overall, the error rate only exceeded 1% at 2500bps for 40kHz and at 3200bps for 25kHz. This is attributed to the transducers’ relatively large Q value, causing ringing at the end of data transfer and increasing bit errors as the bit rate rises. Although not a definitive test of the maximum data rate for ultrasonic through-body transfer, these experiments outline the feasibility of touch-enabled data exchange via SAWs, with adequate performance for transmitting object IDs, metadata, low-frequency sensor data, and URLs. For context, a short URL (136 bits) can be transmitted in approximately 4 ms.

Additionally, we tested several commonly used materials for ultrasonic signal transmission by placing a 40 kHz transducer on surfaces such as granite, wood, plastic, and metal. We also evaluated signal transmission through walls and LCD screens to assess the effectiveness of active data transfer across different environments. All tested media successfully conveyed modulated data. Furthermore, with varying distances between the transducer and the user’s hand wearing HandSAW, the device reliably received encoded data through the surface at ranges from 10 cm to 2 m. This paves the way for developing future interactions and applications, transforming physical objects into digital, intelligent spaces using inaudible modulated signals.

Although HandSAW has demonstrated its capability for transferring modulated signals, the active data transfer mechanism is not intended to challenge the higher speeds of modern wireless technologies like WiFi or Bluetooth. The intent is to transfer small payloads, like UUIDs or URLs, via direct physical interaction. For example, a user can initiate Bluetooth pairing by explicitly touching the target pair of headphones (thus transferring a short data string of 136 bits). Unlike the indiscriminate broadcast of radio packets, this approach gives users control over where their data is transferred via touch, which could be beneficial in scenarios requiring stricter measures for reliable and safe information handling.

6 Applications and Usage

This section highlights real-world applications of HandSAW, showcasing its versatility and potential for future deployment. Key examples include tracking daily activities, such as morning routine and workshop tasks monitoring, and transforming physical spaces into intelligent environments through active, contextually encoded information transfer.

6.1 Daily Activity Recognition

Interactions with objects reflect an individual’s immediate actions. Analyzing the manipulated object or specific hand activity enables systems to infer a user’s current tasks, the broader environmental context, and potential future intentions. For example, opening a refrigerator followed by using a blender may suggest cooking activities. Following Section 5, this section further demonstrated HandSAW’s application scenarios with morning routines and workshop craftsmanship tasks in natural daily settings. HandSAW enables comprehensive recording of daily activities, facilitating precise tracking of an individual’s routine, which is particularly beneficial for long-term health monitoring.

6.1.1 Activity Recognition for Morning Routine. The first scenario examines a common daily activity—the morning routine, which includes personal hygiene and breakfast preparation. Building on the rigorous results from Section 5 with offline ML model training, this section demonstrates HandSAW as a fully integrated, real-time recognition

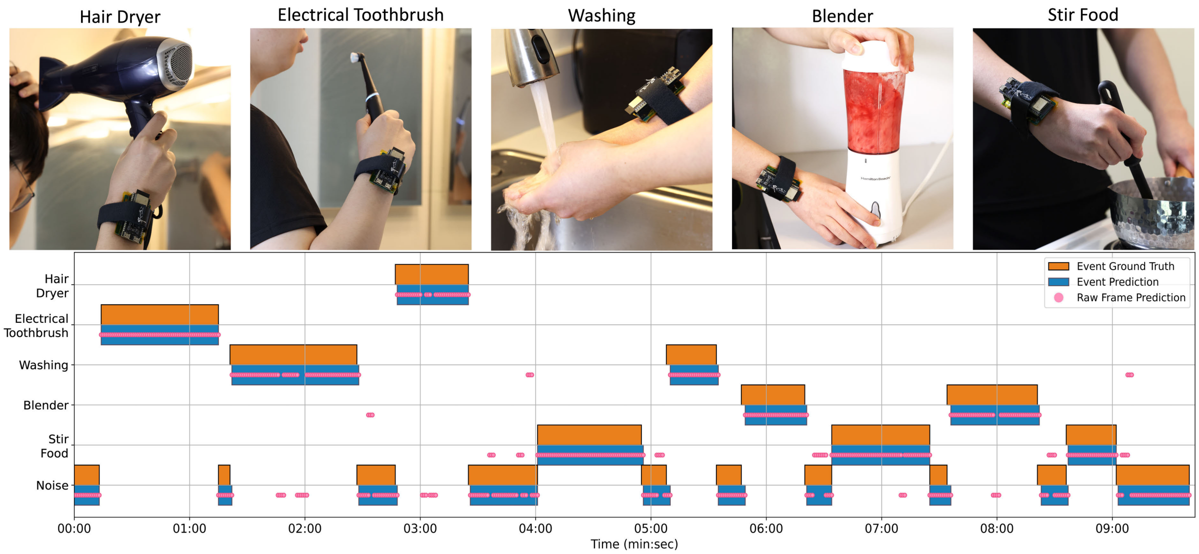

Fig. 16. Morning routine activities include personal hygiene and breakfast preparation. HandSAW detects users' activities and logs predicted events.

system tested with three users in a realistic home environment. Users were asked to simulate their morning routines in a comfortable manner. Suggested hygiene activities included hair drying, using an electric toothbrush, and hand washing. Users brought their own electric toothbrushes, while a hair dryer was provided. We supplied the necessary equipment and ingredients for breakfast, encouraging users to prepare a smoothie and oatmeal porridge. Participants had no specific instructions on how to perform the activities or the sequence of actions and no time restrictions for each task. They were free to talk, play music, or move around the home. Throughout this process, the HandSAW system continuously logged real-time labels of predicted activities, while a researcher observed and manually recorded ground truth labels with timestamps.

Real-time predictions are produced for each individual frame from the current 1-second input data. A laptop recorded the logging labels at approximately four predictions per second, with slight variations due to fluctuations in WiFi connectivity or data processing speed. As illustrated in Figure 16, the raw predictions for each frame are represented by circles. To address incorrect frame predictions, a filter was implemented to exclude overly short predicted frames, thereby enhancing the accuracy of recorded events. Therefore, a predicted user event is generated for each timestamp as shown with bars in Figure 16. The behavior of an example user is illustrated in Figure 16, showing real-time event predictions compared with ground truth events. Results demonstrated HandSAW's capability to accurately track activities with a 100% coverage rate for the event types and no significant latency. The total morning routine activity of one example user lasted 9 minutes and 40 seconds, encompassing 1521 predicted frames. The accuracy of raw predictions for each frame was 89.09%. The accuracy of event prediction reached 97.90% for each timestamp. Intersection over Union (IoU) is calculated by determining the intersection (the maximum start time and the minimum end time between the ground truth and prediction) and the union (the total duration of ground truth and prediction excluding the intersection). The average IoU 0.94 is computed for segments of the same events. Researchers observed diverse behavior patterns among participants, including variations in the action order, duration, and manner of performing activities. Despite these differences, HandSAW consistently logged user activities accurately.

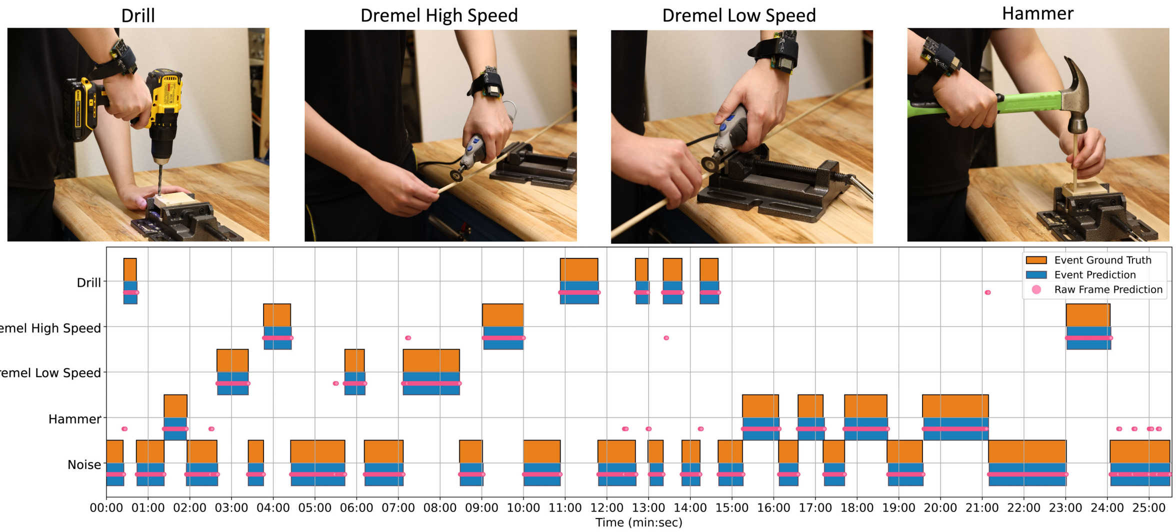

Fig. 17. Craftsmanship activities in the workshop were logged by HandSAW with ground truth and prediction events comparison.

6.1.2 Activity Recognition for Workshop Craftsmanship. In the second scenario, three users were asked to perform craftsmanship tasks in a fabrication workshop using tools such as a drill, a Dremel tool, and a hammer. The workshop was conducted in a fabrication lab with a 3D printer that operated continuously and building air conditioning. For safety, each participant wore safety goggles and first attended a tutorial on safety measures and proper tool use. Each participant was given wooden disks, dowels, and step-by-step instructions for constructing a pen holder. Similar to the morning routine scenario, the HandSAW system captured real-time predicted labels while a researcher recorded ground truth events and timestamps.

As illustrated in Figure 17, the HandSAW system accurately captured all steps of making the pen holder with a 100% coverage rate and no significant latency. The total duration of the user working on a pen holder is 25 minutes and 5 seconds, with 4766 predicted frames. The raw predicted frames achieved a 96.35% accuracy for each timestamp. Frame misclassification primarily occurred when single, short, sharp noises were present, such as the system mistakenly classifying high-amplitude noises as hammers. To address this issue, a filter was implemented to eliminate extremely short events, such as using a drill for only 0.5 seconds, enhancing the precision of the predictions. This adjustment increased the accuracy of predicted events to 98.49%. The slight deviation from 100% accuracy is attributed to minor latency, leading to misclassifications at the beginning and end of activities. The IoU score is 0.98, demonstrating that HandSAW successfully and robustly captured hand-based interactions.

These results validate HandSAW's effectiveness for daily activity monitoring. Its unique approach to detecting skin-propagated acoustic signals preserves user privacy by eliminating the need for cameras or microphones, thereby minimizing the risk of sensitive information exposure. This feature is especially advantageous for health and wellness monitoring, which requires strict privacy standards. Overall, HandSAW's capability to track daily activities provides significant potential for healthcare, providing valuable insights to inform recommendations for healthier habits and ergonomic practices.

Fig. 18. (a) The transducer on a fridge enables HandSAW to deliver grocery reminder messages upon the user touching the fridge. (b) Without requiring transducers or any hardware modifications, a TV can transmit a URL to the HandSAW system using inaudible sound. Users can receive the URL of the image displayed on the TV with just a single touch.

6.2 Smart Space: Context-Driven Information Transfer

Section 5.6 details how HandSAW can be paired with ultrasonic tags to enable data transfer through touch. This ability is particularly useful in providing valuable contextual information about the physical world around users. It transforms numerous purely mechanical objects into channels for digital information in our daily lives. With the intermediary support of HandSAW, information is no longer isolated within each object; instead, it can be connected and transferred under the user's control. A lot of applications can be supported with various types of data transfer. When a user touches the fridge, HandSAW can recognize its object ID and interface with a voice assistant like Alexa, prompting the user, "Grocery Reminder: a dozen of eggs" (Figure 18(a)). Furthermore, since HandSAW supports rich data types such as URLs, it can seamlessly transfer contextual information to companion devices. For example, when a user touches a drinking fountain, the water quality information can be displayed on the user's smartwatch via modulated ultrasonic signals.

In addition to incorporating transducers that facilitate the transmission of ultrasonic signals, common electrical devices also possess the capability to emit distinct signals that are highly characteristic. Through software modifications alone, HandSAW is designed to establish real-time communication with these devices upon user interaction. As shown in Table 2, four of the most commonly used devices in daily life can transmit information such as user IDs, URLs, and authentication data to the HandSAW through either audible or inaudible sounds. For instance, an iPhone emits an 18kHz inaudible sound, encoding information at 100 bps. As AI-generated voices and faces become increasingly prevalent, HandSAW introduces an extra layer of authentication safety, requiring the user's hand and phone to be co-located and in physical contact. Further demonstrating the versatility of HandSAW, Figure 18(b) shows a scenario where a TV or any monitor equipped with a speaker emits an 18kHz inaudible sound. When the user touches the edges of the screen, the URL of the displayed image is transmitted directly to the user's smartphone.

Additionally, vibrations from some devices can induce substantial amplitude sounds, which the VPU can capture through the skin. Leveraging the natural characteristics of game controllers and the iPhone haptics engine, 5Hz vibrations are encoded with information about game scores and user IDs. Subsequently, HandSAW captured and accurately retrieved the data. These enhancements of device functionality require no hardware modifications, only minor software adjustments. More importantly, the modifications do not exceed the devices' original capabilities or functionalities. Through HandSAW system, a ubiquitous intelligent space was created that dynamically facilitates contextual information flow fully controlled by users.

Table 2. Common devices capable of transferring information without any hardware modification.

	MacBook	**iPhone**	**Monitor with Speaker**	**Speaker**	**Game Controller**	**iPhone Haptics**
Bandwidth	audible sounds and inaudible ultrasound	audible sounds and inaudible ultrasound	audible sounds and inaudible ultrasound	audible sounds and inaudible ultrasound	<5hz vibration	<5hz vibration
Application Examples	URL transfer, text transfer, user ID transfer, authentication	URL transfer, text transfer, user ID transfer, authentication	URL transfer, text transfer, user ID transfer, authentication	URL transfer, text transfer, user ID transfer, authentication	Sending game score, player ID, gaming state	user ID transfer, authentication, information-encoded notification

7 Discussion

This section contextualizes HandSAW by comparing it with other sensing techniques. Additionally, several discussion points are presented, including advanced machine learning methods for reducing data collection, as well as an examination of current limitations and potential future work.

7.1 Comparison with Other Sensing Techniques

HandSAW demonstrated broad applicability and reliable performance in recognizing hand-based interactions across diverse users and environments. A comparison of several standard sensing techniques is listed in Table 3 for better contextualization of this research. Examples of the most commonly used sensing mechanisms include IMU, camera, microphone, and magnetic sensors. Previous systems often require extensive instrumentation of the user, are limited to specific types of devices, lack generalizability to unseen users, or have inadequate real-time noise rejection capabilities. For example, magnetic-based or coil-based sensing techniques are limited to applications involving electrical devices or those that generate a magnetic field [65, 66, 110]. Compared to other acoustic-based sensing systems [5, 54, 63, 113], HandSAW offers a wide frequency range, capturing rich acoustic signatures across both the audible and inaudible spectrum. Moreover, HandSAW can effectively eliminate airborne sounds, which enables the rejection of environmental speech and other noises. Additionally, generalizability is important, as collecting data from every new user before system use is impractical. The uniqueness of the acoustic features captured by HandSAW is that the frequency distribution for a specific class remains consistent across different users. Maintaining a 91.6% accuracy with a new, unseen user, HandSAW demonstrated robust performance. It achieved the highest cross-user accuracy among the systems compared here. Moreover, HandSAW has demonstrated feasibility for active sensing, unlocking various application scenarios. Viband [52]reports achieving a data rate of 400 bps using 16-QAM with a 300 Hz (audible) carrier and a 7-16% error rate. By contrast, HandSAW enables transfer data at 3200 bps with a 25 kHz (inaudible) carrier and a 1% error rate. HandSAW has also demonstrated robust data transfer capabilities with longer distances. HandSAW demonstrated the ability to instrument an entire countertop and detect data from 2 meters away, whereas Viband requires direct contact with the transmitter. Furthermore, real-time, onboard ML can increase the practicality of this system in real-world scenarios. Compared to previous approaches [5, 21, 113], HandSAW stands out by offering the robustness and efficiency necessary for real-world deployment, making it a viable solution for real-world applications. HandSAW achieves state-of-the-art performance comparable to other systems listed here, while only six prior works support real-time predictions, and none offer onboard machine learning inference capabilities.

7.2 Limitations and Future Work

In this section, we describe current limitations of the HandSAW system and avenues for future work to address them.

Table 3. Comparison of other typical sensing techniques in the literature. Each paper is represented by either the project name or the first author's last name.

	Technique	Audio Range	Performance	Active Sensing	Noise Rejection	Real-time	On-board ML	Cross-user Accuracy
OSense [8]	IMU	-	10 hand activities 87%	✗	✗	✗	✗	✗
ViBand [52]	IMU	-	29 object recognition 91.5%	✓	✓	✓	✗	17 users, 91.5%
Laput [51]	IMU	-	25 hand activities 95.2%	✗	✗	✓	✗	12 users, 79.2%
VibEye [80]	actuator, accelerometer	-	16 materials (92.5%) 25 objects recognition	✗	✓	✗	✗	✗
WristSens [67]	accelerometer camera	-	10 home activities not mentioned	✗	✗	✓	✗	✗
Maekawa [65]	magnetic sensor	-	14 home activities over 75%	✗	✗	✗	✗	✗
MagnifiSense [110]	magneto-inductive coil	-	12 home devices 83.9%	✗	✓	✗	✗	✗
Maekawa [66]	coil of wire	-	14 electrical devices 84%	✗	✗	✗	✗	5 users, 82.2%
Ohnishi [81]	camera	-	23 ADLs 89.7%	✗	✗	✓	✗	✗
Fan [21]	EMG sensors	-	15 object recognition 82.5%	✗	✗	✗	✗	12 users, 47.9%
Z-Ring [108]	electrodes on the ring	-	5 hand activities 93%	✗	✗	✓	✗	✗
EchoWrist [54]	microphone, speaker	18-21kHz or 20-24kHz	12 object recognition, 97.6%	✓	✓	✓	✗	12 users, 82.8%
Lukowicz [63]	accelerometer, microphone	2khz	8 workshop activities 83.5%	✗	✗	✗	✗	✗
Ward [113]	accelerometer, microphone	2khz	21 workshop activities, 98%	✗	✗	✗	✗	5 users, 87%
Bhattacharya [5]	IMU, microphone	22.05KHz	23 home activities 94.3%	✗	✗	✗	✗	15 users, 89.7%(F1)
HandSAW	SAWs sensor	0-48kHz	20 hand-based activities, 99%	✓	✓	✓	✓	16 users, 91.6%

7.2.1 Advanced ML Methods for Reducing Data Collection. While HandSAW achieved a 91.6% accuracy on test-train splits, even with unseen users, collecting labelled training data for new classes of objects presents scalability challenges that future work should address. A promising direction is few-shot learning, which identifies underlying patterns in data with only a limited number of training samples. Unlike traditional deep learning models, which require extensive datasets, few-shot learning emphasizes 'learning how to learn.' This is achieved by training models on diverse objects, encouraging the system to recognize essential features and patterns across classes, and enabling it to generalize effectively from minimal labelled examples. This approach is particularly advantageous because many objects produce distinct frequencies of mechanical vibrations based on their operation, independent of how users hold these devices. This independence means that custom models for individual users aren't necessary; instead, models can be trained for the objects themselves. As a result, leveraging the robust feature extraction and association capabilities of few-shot learning can significantly reduce the burden of data collection from users. In addition to few-shot learning, generative learning and diffusion models offer promising avenues for minimizing the need for extensive training data.

7.2.2 False Positives Misclassification & Noises. Several design aspects have been implemented to mitigate false positive situations, especially caused by single, short, and noisy events. For activity recognition, HandSAW operates at a 48kHz sampling rate. The system uses a 1-second sliding window with a 0.25-second step size to enhance responsiveness and prevent event segmentation across multiple windows. For each 1-second data frame, a custom MFCC feature extraction algorithm captures features across the 0Hz to 24kHz. Although impulse events are brief in the time domain, they contain rich frequency characteristics that aid in accurate classification. We recognize, however, that this approach could potentially miss very short events. In such cases, reducing the window size would be necessary. The results show that HandSAW effectively classified typical length user object interaction events, including short events like finger taps and claps. Moreover, the filter introduced in Section 6 allows the rejection of false predictions by effectively "low-pass filtering" the data, demonstrating that HandSAW is robust in detecting user events. However, better algorithms can be developed to further address, mitigate, and overcome this issue.

7.2.3 Contact Conditions with the Wrist. One constraint of HandSAW is that the VPU must maintain contact with the user's skin to detect valid signals. The current implementation with four VPUs increases the likelihood that at least one VPU will maintain contact with the user's skin, enhancing signal detection reliability. The results in Section 5 demonstrate that using a single VPU yields accuracy that is only marginally lower compared to using four VPUs. Despite this, the current prototype is not optimized for user comfort or maintaining consistent contact with the user's skin. Future design improvements could focus on developing a wristband that better conforms to the human body and its topography. Incorporating soft, flexible, springy, and conformal materials may also significantly enhance both comfort and usability.

8 Conclusion

This work presents HandSAW, a fully wireless system capable of passively recognizing 20 events with >99% per user accuracy and 91.6% accuracy across users, demonstrating robust performance even for a new user without any calibration data. A Gini impurity analysis shows that the sensors used in HandSAW are position invariant, allowing significant flexibility in sensor placement for future designs. HandSAW also supports active data transfer through 40kHz ASK, sustaining 3000bps with about 1% bit error. To demonstrate the utility of these capabilities, two illustrative applications are developed: wearable tracking of activities of daily living and active data transfer with modulated ultrasonic signals upon touch. We believe these findings spur greater interest in on-body applications of SAW-based sensing and enable various applications in the future.

References

[1] Bilel Ben Atitallah, Muhammed Bilal Abbasi, Rim Barioul, Dhouha Bouchaala, Nabil Derbel, and Olfa Kanoun. 2020. Simultaneous pressure sensors monitoring system for hand gestures recognition. In *2020 IEEE SENSORS*. IEEE, 1–4.

[2] Joonsung Bae, Hyunwoo Cho, Kiseok Song, Hyungwoo Lee, and Hoi-Jun Yoo. 2012. The signal transmission mechanism on the surface of human body for body channel communication. *IEEE Transactions on microwave theory and techniques* 60, 3 (2012), 582–593.

[3] Lu Bai, Chris Yeung, Christos Efstratiou, and Moyra Chikomo. 2019. Motion2Vector: Unsupervised learning in human activity recognition using wrist-sensing data. In *Adjunct proceedings of the 2019 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2019 ACM international symposium on wearable computers*. 537–542.

[4] Behringer. 2023. Behringer | Product | UM2. https://www.behringer.com/product.html?modelCode=0805-AAL.

[5] Sarnab Bhattacharya, Rebecca Adaimi, and Edison Thomaz. 2022. Leveraging Sound and Wrist Motion to Detect Activities of Daily Living with Commodity Smartwatches. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 42 (jul 2022), 28 pages. doi:10.1145/3534582

[6] Kanad K Biswas and Saurav Kumar Basu. 2011. Gesture recognition using microsoft kinect®. In *The 5th international conference on automation, robotics and applications*. IEEE, 100–103.

[7] Andreas Braun, Stefan Krepp, and Arjan Kuijper. 2015. Acoustic tracking of hand activities on surfaces. In *Proceedings of the 2nd international Workshop on Sensor-based Activity Recognition and Interaction*. 1–5.

[8] Thisum Buddhika, Haimo Zhang, Chamod Weerasinghe, Suranga Nanayakkara, and Roger Zimmermann. 2019. OSense: Object-activity Identification Based on Gasping Posture and Motion. In *Proceedings of the 10th Augmented Human International Conference 2019* (Reims, France) *(AH2019)*. Association for Computing Machinery, New York, NY, USA, Article 13, 5 pages. doi:10.1145/3311823.3311841

[9] Michael Buettner, Richa Prasad, Matthai Philipose, and David Wetherall. 2009. Recognizing Daily Activities with RFID-Based Sensors. In *Proceedings of the 11th International Conference on Ubiquitous Computing* (Orlando, Florida, USA) *(UbiComp '09)*. Association for Computing Machinery, New York, NY, USA, 51–60. doi:10.1145/1620545.1620553

[10] Shumin Cao, Xin He, Peide Zhu, Mingshi Chen, Xiangyang Li, and Panlong Yang. 2018. IPand: accurate gesture input with ambient acoustic sensing on hand. In *2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC)*. IEEE, 1–8.

[11] Yetong Cao, Fan Li, Huijie Chen, Xiaochen Liu, Chunhui Duan, and Yu Wang. 2023. I Can Hear You Without a Microphone: Live Speech Eavesdropping From Earphone Motion Sensors. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*. IEEE, 1–10.

[12] Kun Chen, Xing Fu, Dante J Dorantes-Gonzalez, Zimo Lu, Tingting Li, Yanning Li, Sen Wu, and Xiaotang Hu. 2014. Simulation study of melanoma detection in human skin tissues by laser-generated surface acoustic waves. *Journal of Biomedical Optics* 19, 7 (2014), 077007–077007.

[13] Mingshi Chen, Panlong Yang, Shumin Cao, Maotian Zhang, and Ping Li. 2018. WritePad: Consecutive Number Writing on Your Hand With Smart Acoustic Sensing. *IEEE Access* 6 (2018), 77240–77249. doi:10.1109/ACCESS.2018.2880980

[14] Wenqiang Chen, Shupei Lin, Zhencan Peng, Farshid Salemi Parizi, Seongkook Heo, Shwetak Patel, Wojciech Matusik, Wei Zhao, and John Stankovic. 2024. ViObject: Harness Passive Vibrations for Daily Object Recognition with Commodity Smartwatches. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 1, Article 5 (mar 2024), 26 pages. doi:10.1145/3643547

[15] Yuqing Chen and Yang Xue. 2015. A deep learning approach to human activity recognition based on single accelerometer. In *2015 IEEE international conference on systems, man, and cybernetics*. IEEE, 1488–1492.

[16] Gabe Cohn, Daniel Morris, Shwetak N Patel, and Desney S Tan. 2011. Your noise is my command: sensing gestures using the body as an antenna. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 791–800.

[17] Artem Dementyev and Joseph A. Paradiso. 2014. WristFlex: Low-Power Gesture Input with Wrist-Worn Pressure Sensors. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 161–166. doi:10.1145/2642918.2647396

[18] DigiKey. 2023. ADXL1001 Accelerometer. https://www.digikey.com/en/products/detail/analog-devices-inc/ADXL1001BCPZ-RL/7203662.

[19] Murata Electronics. 2023. 7BB-27-4L0 Murata Electronics | Audio Products | DigiKey. https://www.digikey.com/en/products/detail/murata-electronics/7BB-27-4L0/4358154.

[20] ESP32. 2023. Inter-IC Sound (I2S) - ESP32 - — ESP-IDF Programming Guide latest documentation. https://docs.espressif.com/projects/esp-idf/en/latest/esp32/api-reference/peripherals/i2s.html.

[21] Junjun Fan, Xiangmin Fan, Feng Tian, Yang Li, Zitao Liu, Wei Sun, and Hongan Wang. 2018. What is That in Your Hand? Recognizing Grasped Objects via Forearm Electromyography Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 161 (dec 2018), 24 pages. doi:10.1145/3287039

[22] Camelia Gabriel, Sami Gabriel, and YE Corthout. 1996. The dielectric properties of biological tissues: I. Literature survey. *Physics in medicine & biology* 41, 11 (1996), 2231.

[23] Sami Gabriel, RW Lau, and Camelia Gabriel. 1996. The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz. *Physics in medicine & biology* 41, 11 (1996), 2251.

[24] Valerie Galluzzi, Ted Herman, and Philip Polgreen. 2015. Hand Hygiene Duration and Technique Recognition Using Wrist-Worn Sensors. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks* (Seattle, Washington) *(IPSN '15)*. Association for Computing Machinery, New York, NY, USA, 106–117. doi:10.1145/2737095.2737106

[25] Yan Gao, Yang Long, Yu Guan, Anna Basu, Jessica Baggaley, and Thomas Plötz. 2020. Automated general movement assessment for perinatal stroke screening in infants. *Smart Assisted Living: Toward An Open Smart-Home Infrastructure* (2020), 167–187.

[26] Asma Ghandeharioun, Szymon Fedor, Lisa Sangermano, Dawn Ionescu, Jonathan Alpert, Chelsea Dale, David Sontag, and Rosalind Picard. 2017. Objective assessment of depressive symptoms with machine learning and wearable sensors data. In *2017 seventh international conference on affective computing and intelligent interaction (ACII)*. IEEE, 325–332.

[27] Giacomo Giorgi, Fabio Martinelli, Andrea Saracino, and Mina Sheikhalishahi. 2017. Try walking in my shoes, if you can: Accurate gait recognition through deep learning. In *Computer Safety, Reliability, and Security: SAFECOMP 2017 Workshops, ASSURE, DECSoS, SASSUR, TELERISE, and TIPS, Trento, Italy, September 12, 2017, Proceedings 36*. Springer, 384–395.

[28] Mohit Goyal, Sahil Modi, Rishabh Goyal, and Saurabh Gupta. 2022. Human hands as probes for interactive object understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3293–3303.

[29] Connor Griffin and Victor Giurgiutiu. 2023. Piezoelectric wafer active sensor transducers for acoustic emission applications. *Sensors* 23, 16 (2023), 7103.

[30] Yan He, Hanyan Zhang, Edwin Yang, and Song Fang. 2020. Virtual Step PIN Pad: Towards Foot-input Authentication Using Geophones. In *2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*. IEEE, 649–657.

[31] Yu-Jin Hong, Ig-Jae Kim, Sang Chul Ahn, and Hyoung-Gon Kim. 2010. Mobile health monitoring system based on activity recognition using accelerometer. *Simulation Modelling Practice and Theory* 18, 4 (2010), 446–455.

[32] Fang Hu, Peng He, Songlin Xu, Yin Li, and Cheng Zhang. 2020. FingerTrak: Continuous 3D Hand Pose Tracking by Deep Learning Hand Silhouettes Captured by Miniature Thermal Cameras on Wrist. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 71 (jun 2020), 24 pages. doi:10.1145/3397306

[33] Inc. Input/Output. [n. d.]. Piezo Element. https://www.sparkfun.com/products/10293.

[34] InvenSense. 2023. MPU-9250-Register-Map. https://cdn.sparkfun.com/assets/learn_tutorials/5/5/0/MPU-9250-Register-Map.pdf.

[35] Yasha Iravantchi, Karan Ahuja, Mayank Goel, Chris Harrison, and Alanson Sample. 2021. PrivacyMic: Utilizing Inaudible Frequencies for Privacy Preserving Daily Activity Recognition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 198, 13 pages. doi:10.1145/3411764.3445169

[36] Yasha Iravantchi, Mayank Goel, and Chris Harrison. 2019. BeamBand: Hand Gesture Sensing with Ultrasonic Beamforming. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3290605.3300245

[37] Yasha Iravantchi, Yang Zhang, Evi Bernitsas, Mayank Goel, and Chris Harrison. 2019. Interferi: Gesture Sensing Using On-Body Acoustic Interferometry. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300506

[38] Yasha Iravantchi, Yi Zhao, Kenrick Kin, and Alanson P. Sample. 2023. SAWSense: Using Surface Acoustic Waves for Surface-Bound Event Recognition. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 422, 18 pages. doi:10.1145/3544548.3580991

[39] Deming Jiang, Jingwen Liu, Yuxiang Pan, Liujing Zhuang, and Ping Wang. 2021. Surface acoustic wave (SAW) techniques in tissue engineering. *Cell and Tissue Research* 386, 2 (2021), 215–226.

[40] Shuo Jiang, Bo Lv, Weichao Guo, Chao Zhang, Haitao Wang, Xinjun Sheng, and Peter B. Shull. 2018. Feasibility of Wrist-Worn, Real-Time Hand, and Surface Gesture Recognition via sEMG and IMU Sensing. *IEEE Transactions on Industrial Informatics* 14, 8 (2018), 3376–3385. doi:10.1109/TII.2017.2779814

[41] Shuo Jiang, Bo Lv, Xinjun Sheng, Chao Zhang, Haitao Wang, and Peter B Shull. 2016. Development of a real-time hand gesture recognition wristband based on sEMG and IMU sensing. In *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 1256–1261.

[42] Yujian Jiang, Lin Song, Junming Zhang, Yang Song, and Ming Yan. 2022. Multi-Category Gesture Recognition Modeling Based on sEMG and IMU Signals. *Sensors* 22 (08 2022), 5855. doi:10.3390/s22155855

[43] Vimal Kakaraparthi, Qijia Shao, Charles J. Carver, Tien Pham, Nam Bui, Phuc Nguyen, Xia Zhou, and Tam Vu. 2021. FaceSense: Sensing Face Touch with an Ear-worn System. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 110 (sep 2021), 27 pages. doi:10.1145/3478129

[44] Haik Kalantarian, Nabil Alshurafa, Tuan Le, and Majid Sarrafzadeh. 2015. Monitoring eating habits using a piezoelectric sensor-based necklace. *Computers in biology and medicine* 58 (2015), 46–55.

[45] Aleksander A Karabutov. 1985. Laser excitation of surface acoustic waves: a new direction in opto-acoustic spectroscopy of a solid. *Soviet Physics Uspekhi* 28, 11 (nov 1985), 1042. doi:10.1070/PU1985v028n11ABEH003981

[46] Hojoong Kim, Young-Tae Kwon, Hyo-Ryoung Lim, Jong-Hoon Kim, Yun-Soung Kim, and Woon-Hong Yeo. 2021. Recent advances in wearable sensors and integrated functional devices for virtual and augmented reality applications. *Advanced Functional Materials* 31, 39 (2021), 2005692.

[47] Jungsoo Kim, Jiasheng He, Kent Lyons, and Thad Starner. 2007. The gesture watch: A wireless contact-free gesture based wrist interface. In *2007 11th IEEE International Symposium on Wearable Computers*. IEEE, 15–22.

[48] Jeong-Kyun Kim, Da-Som Oh, Kangbok Lee, and Sang Gi Hong. 2022. Fall Detection Based on Interpretation of Important Features with Wrist-Wearable Sensors. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking* (Sydney, NSW, Australia) *(MobiCom '22)*. Association for Computing Machinery, New York, NY, USA, 823–825. doi:10.1145/3495243.3558250

[49] Ekaterina Kutafina, David Laukamp, Ralf Bettermann, Ulrik Schroeder, and Stephan M Jonas. 2016. Wearable sensors for eLearning of manual tasks: using forearm EMG in hand hygiene training. *Sensors* 16, 8 (2016), 1221.

[50] Nicholas D. Lane, Petko Georgiev, Cecilia Mascolo, and Ying Gao. 2015. ZOE: A Cloud-Less Dialog-Enabled Continuous Sensing Wearable Exploiting Heterogeneous Computation. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (Florence, Italy) *(MobiSys '15)*. Association for Computing Machinery, New York, NY, USA, 273–286. doi:10.1145/2742647.2742672

[51] Gierad Laput and Chris Harrison. 2019. Sensing Fine-Grained Hand Activity with Smartwatches. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300568

[52] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 321–333. doi:10.1145/2984511.2984582

[53] Gierad Laput, Chouchang Yang, Robert Xiao, Alanson Sample, and Chris Harrison. 2015. EM-Sense: Touch Recognition of Uninstrumented, Electrical and Electromechanical Objects. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 157–166. doi:10.1145/2807442.2807481

[54] Chi-Jung Lee, Ruidong Zhang, Devansh Agarwal, Tianhong Catherine Yu, Vipin Gunda, Oliver Lopez, James Kim, Sicheng Yin, Boao Dong, Ke Li, Mose Sakashita, Francois Guimbretiere, and Cheng Zhang. 2024. EchoWrist: Continuous Hand Pose Tracking and

Hand-Object Interaction Recognition Using Low-Power Active Acoustic Sensing On a Wristband. arXiv:2401.17409 [cs.HC]
[55] Seungyon Claire Lee, Bohao Li, and Thad Starner. 2011. AirTouch: Synchronizing in-air hand gesture and on-body tactile feedback to augment mobile gesture interaction. In *2011 15th Annual International Symposium on Wearable Computers*. IEEE, 3–10.
[56] C Li, G Guan, R Wang, and Z Huang. 2016. Mechanical characterization of skin using surface acoustic waves. In *Imaging in Dermatology*. Elsevier, 327–340.
[57] Hanchuan Li, Can Ye, and Alanson P. Sample. 2015. IDSense: A Human Object Interaction Detection System Based on Passive UHF RFID. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2555–2564. doi:10.1145/2702123.2702178
[58] Dawei Liang, Guihong Li, Rebecca Adaimi, Radu Marculescu, and Edison Thomaz. 2022. AudioIMU: Enhancing Inertial Sensing-Based Activity Recognition with Acoustic Models. In *Proceedings of the 2022 ACM International Symposium on Wearable Computers* (Cambridge, United Kingdom) *(ISWC '22)*. Association for Computing Machinery, New York, NY, USA, 44–48. doi:10.1145/3544794.3558471
[59] librosa. 2023. librosa.feature.mfcc — librosa 0.10.1 documentation. https://librosa.org/doc/main/generated/librosa.feature.mfcc.html.
[60] Kang Ling, Haipeng Dai, Yuntang Liu, and Alex X. Liu. 2018. UltraGesture: Fine-Grained Gesture Sensing and Recognition. In *2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)* (Hong Kong, China). IEEE Press, 1–9. doi:10.1109/SAHCN.2018.8397099
[61] Donghan Liu, Dinghuang Zhang, and Honghai Liu. 2022. Dynamic Hand Gesture Recognition For Numeral Handwritten Via A-Mode Ultrasound. In *Intelligent Robotics and Applications: 15th International Conference, ICIRA 2022, Harbin, China, August 1–3, 2022, Proceedings, Part II* (Harbin, China). Springer-Verlag, Berlin, Heidelberg, 614–625. doi:10.1007/978-3-031-13822-5_55
[62] LSM9DS1. 2023. FemmeVerbeek/Arduino_LSM9DS1: LSM9DS1 Library for Arduino. https://github.com/FemmeVerbeek/Arduino_LSM9DS1.
[63] Paul Lukowicz, Jamie A Ward, Holger Junker, Mathias Stäger, Gerhard Tröster, Amin Atrash, and Thad Starner. 2004. Recognizing workshop activity using body worn microphones and accelerometers. In *Pervasive Computing: Second International Conference, PERVASIVE 2004, Linz/Vienna, Austria, April 21-23, 2004. Proceedings 2*. Springer, 18–32.
[64] Gan Luo, Panlong Yang, Mingshi Chen, and Ping Li. 2020. HCI on the Table: Robust Gesture Recognition Using Acoustic Sensing in Your Hand. *IEEE Access* 8 (2020), 31481–31498. doi:10.1109/ACCESS.2020.2973305
[65] Takuya Maekawa, Yasue Kishino, Yasushi Sakurai, and Takayuki Suyama. 2011. Recognizing the Use of Portable Electrical Devices with Hand-Worn Magnetic Sensors. 276–293. doi:10.1007/978-3-642-21726-5_18
[66] Takuya Maekawa, Yasue Kishino, Yutaka Yanagisawa, and Yasushi Sakurai. 2012. Recognizing Handheld Electrical Device Usage with Hand-Worn Coil of Wire. 234–252. doi:10.1007/978-3-642-31205-2_15
[67] Takuya Maekawa, Yasue Kishino, Yutaka Yanagisawa, and Yasushi Sakurai. 2012. WristSense: Wrist-worn sensor device with camera for daily activity recognition. In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*. 510–512. doi:10.1109/PerComW.2012.6197551
[68] Takuya Maekawa, Yutaka Yanagisawa, Yasue Kishino, Katsuhiko Ishiguro, Koji Kamei, Yasushi Sakurai, and Takeshi Okadome. 2010. Object-Based Activity Recognition with Heterogeneous Sensors on Wrist. In *Proceedings of the 8th International Conference on Pervasive Computing* (Helsinki, Finland) *(Pervasive'10)*. Springer-Verlag, Berlin, Heidelberg, 246–264. doi:10.1007/978-3-642-12654-3_15
[69] Oleg Makaussov, Mikhail Krassavin, Maxim Zhabinets, and Siamac Fazli. 2020. A low-cost, IMU-based real-time on device gesture recognition glove. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 3346–3351.
[70] Sneha Mane, Ruta Kambli, Faruk Kazi, and Navdeep Singh. 2015. Hand Motion Recognition from Single Channel Surface EMG Using Wavelet & Artificial Neural Network. *Procedia Computer Science* 49 (12 2015), 58–65. doi:10.1016/j.procs.2015.04.227
[71] Alexei A Maradudin and GI Stegeman. 1991. Surface acoustic waves. In *Surface Phonons*. Springer, 5–35.
[72] Jess McIntosh, Asier Marzo, Mike Fraser, and Carol Phillips. 2017. EchoFlex: Hand Gesture Recognition Using Ultrasound Imaging. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1923–1934. doi:10.1145/3025453.3025807
[73] Joan Melià-Seguí and Rafael Pous. 2014. Human-object interaction reasoning using RFID-enabled smart shelf. In *2014 International Conference on the Internet of Things (IOT)*. IEEE, 37–42.
[74] Karan D Mohan and Amy L Oldenburg. 2012. Elastography of soft materials and tissues by holographic imaging of surface acoustic waves. *Optics express* 20, 17 (2012), 18887–18897.
[75] Shreyash Mohatta, Ramakrishna Perla, Gaurav Gupta, Ehtesham Hassan, and Ramya Hebbalaguppe. 2017. Robust hand gestural interaction for smartphone based AR/VR applications. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 330–335.
[76] Adiyan Mujibiya, Xiang Cao, Desney S Tan, Dan Morris, Shwetak N 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*. 189–198.
[77] Chaithanya Mummadi, Frederic Philips, Keshav Verma, Shivaji Kasireddy, Philipp Scholl, Jochen Kempfle, and Kristof Van Laerhoven. 2018. Real-Time and Embedded Detection of Hand Gestures with an IMU-Based Glove. *Informatics* 5 (06 2018), 28. doi:10.3390/

informatics5020028
[78] Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P Patil. 2003. LANDMARC: Indoor location sensing using active RFID. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003)*. IEEE, 407–415.
[79] Jon Nordby, Mark Cooke, and Adam Horvath. 2019. emlearn: Machine Learning inference engine for Microcontrollers and Embedded Devices. doi:10.5281/zenodo.2589394
[80] Seungjae Oh, Gyeore Yun, Chaeyong Park, Jinsoo Kim, and Seungmoon Choi. 2019. VibEye: Vibration-Mediated Object Recognition for Tangible Interactive Applications. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300906
[81] Katsunori Ohnishi, Atsushi Kanehira, Asako Kanezaki, and Tatsuya Harada. 2016. Recognizing activities of daily living with a wrist-mounted camera. (2016), 3103–3111.
[82] Brice Parilusyan, Marc Teyssier, Valentin Martinez-Missir, Clément Duhart, and Marcos Serrano. 2022. Sensurfaces: A Novel Approach for Embedded Touch Sensing on Everyday Surfaces. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 67 (jul 2022), 19 pages. doi:10.1145/3534616
[83] Siddika Parlak and Ivan Marsic. 2013. Detecting Object Motion Using Passive RFID: A Trauma Resuscitation Case Study. *IEEE Transactions on Instrumentation and Measurement* 62, 9 (2013), 2430–2437. doi:10.1109/tim.2013.2258772
[84] Matthai Philipose, Kenneth P Fishkin, Mike Perkowitz, Donald J Patterson, Dieter Fox, Henry Kautz, and Dirk Hahnel. 2004. Inferring activities from interactions with objects. *IEEE pervasive computing* 3, 4 (2004), 50–57.
[85] Lord Rayleigh. 1885. On waves propagated along the plane surface of an elastic solid. *Proceedings of the London mathematical Society* 1, 1 (1885), 4–11.
[86] Patricio Rivera, Edwin Valarezo, and Tae-Seong Kim. 2021. An Integrated ARMA-Based Deep Autoencoder and GRU Classifier System for Enhanced Recognition of Daily Hand Activities. *International Journal of Pattern Recognition and Artificial Intelligence* 35, 06 (2021), 2152006. doi:10.1142/S0218001421520066 arXiv:https://doi.org/10.1142/S0218001421520066
[87] Patricio Rivera, Edwin Valarezo Añazco, Mun-Taek Choi, and Tae-Seong Kim. 2017. Recognition of Human Hand Activities Based on a Single Wrist IMU Using Recurrent Neural Networks. *International Journal of Pharma Medicine and Biological Sciences* 6 (01 2017), 114–118. doi:10.18178/ijpmbs.6.4.114-118
[88] Yu Sang, Laixi Shi, and Yimin Liu. 2018. Micro Hand Gesture Recognition System Using Ultrasonic Active Sensing. *IEEE Access* 6 (2018), 49339–49347. doi:10.1109/ACCESS.2018.2868268
[89] G. Enrico Santagati and Tommaso Melodia. 2015. U-Wear: Software-Defined Ultrasonic Networking for Wearable Devices. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (Florence, Italy) *(MobiSys '15)*. Association for Computing Machinery, New York, NY, USA, 241–256. doi:10.1145/2742647.2742655
[90] Philipp Schilk, Niccolò Polvani, Andrea Ronco, Milos Cernak, and Michele Magno. 2023. In-Ear-Voice: Towards Milli-Watt Audio Enhancement With Bone-Conduction Microphones for In-Ear Sensing Platforms. In *Proceedings of the 8th ACM/IEEE Conference on Internet of Things Design and Implementation* (San Antonio, TX, USA) *(IoTDI '23)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3576842.3582365
[91] Dieter Schneider. 2019. Laser-induced surface acoustic waves for material testing. (2019).
[92] Breawn Schoun, Hawkar Oagaz, and Min-Hyung Choi. 2021. Corner-based square fiducial marker detection for hand-manipulated AR objects. In *2021 IEEE International Conference on Intelligent Reality (ICIR)*. IEEE, 31–38.
[93] Dandan Shan, Jiaqi Geng, Michelle Shu, and David F Fouhey. 2020. Understanding human hands in contact at internet scale. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9869–9878.
[94] Dandan Shan, Richard Higgins, and David Fouhey. 2021. COHESIV: Contrastive object and hand embedding segmentation in video. *Advances in Neural Information Processing Systems* 34 (2021), 5898–5909.
[95] Yunzhou Shi, Megan Manco, Dominique Moyal, Gil Huppert, Hitoshi Araki, Anthony Banks, Hemant Joshi, Richard McKenzie, Alex Seewald, Guy Griffin, et al. 2018. Soft, stretchable, epidermal sensor with integrated electronics and photochemistry for measuring personal UV exposures. *PloS one* 13, 1 (2018), e0190233.
[96] Yilei Shi, Haimo Zhang, Kaixing Zhao, Jiashuo Cao, Mengmeng Sun, and Suranga Nanayakkara. 2020. Ready, Steady, Touch! Sensing Physical Contact with a Finger-Mounted IMU. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 59 (jun 2020), 25 pages. doi:10.1145/3397309
[97] Xinyu Song, Shirdi Shankara van de Ven, Shugeng Chen, Peiqi Kang, Qinghua Gao, Jie Jia, and Peter B. Shull. 2022. Proposal of a Wearable Multimodal Sensing-Based Serious Games Approach for Hand Movement Training After Stroke. *Frontiers in Physiology* 13 (2022), 811950. doi:10.3389/fphys.2022.811950
[98] Yale Song, David Demirdjian, and Randall Davis. 2011. Tracking body and hands for gesture recognition: Natops aircraft handling signals database. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 500–506.
[99] Sonion. [n. d.]. Sonion • Improving people's quality of life. https://www.sonion.com/hearing/bone-conduction-sensors-and-actuators/vpu-voice-pick-up-sensor/.

[100] Maja Stikic, Tâm Huynh, Kristof Van Laerhoven, and Bernt Schiele. 2008. ADL recognition based on the combination of RFID and accelerometer sensing. In *2008 second international conference on pervasive computing technologies for healthcare*. IEEE, 258–263.

[101] Jannis Strecker, Khakim Akhunov, Federico Carbone, Kimberly García, Kenan Bektaş, Andres Gomez, Simon Mayer, and Kasim Sinan Yildirim. 2023. MR Object Identification and Interaction: Fusing Object Situation Information from Heterogeneous Sources. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 3, Article 124 (sep 2023), 26 pages. doi:10.1145/3610879

[102] Subramanian Sundaram, Petr Kellnhofer, Yunzhu Li, Jun-Yan Zhu, Antonio Torralba, and Wojciech Matusik. 2019. Learning the signatures of the human grasp using a scalable tactile glove. *Nature* 569, 7758 (2019), 698–702.

[103] Hoang Truong, Shuo Zhang, Ufuk Muncuk, Phuc Nguyen, Nam Bui, Anh Nguyen, Qin Lv, Kaushik Chowdhury, Thang Dinh, and Tam Vu. 2018. CapBand: Battery-Free Successive Capacitance Sensing Wristband for Hand Gesture Recognition. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems* (Shenzhen, China) *(SenSys '18)*. Association for Computing Machinery, New York, NY, USA, 54–67. doi:10.1145/3274783.3274854

[104] Ayumu Tsuboi, Mamoru Hirota, Junki Sato, Masayuki Yokoyama, and Masao Yanagisawa. 2017. A Proposal for Wearable Controller Device and Finger Gesture Recognition Using Surface Electromyography. In *SIGGRAPH Asia 2017 Posters* (Bangkok, Thailand) *(SA '17)*. Association for Computing Machinery, New York, NY, USA, Article 9, 2 pages. doi:10.1145/3145690.3145731

[105] Yu-Chih Tung, Duc Bui, and Kang G. Shin. 2018. Cross-Platform Support for Rapid Development of Mobile Acoustic Sensing Applications. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services* (Munich, Germany) *(MobiSys '18)*. Association for Computing Machinery, New York, NY, USA, 455–467. doi:10.1145/3210240.3210312

[106] Edwin Valarezo Añazco, Seung Ju Han, Kangil Kim, Patricio Rivera Lopez, Tae-Seong Kim, and Sangmin Lee. 2021. Hand gesture recognition using single patchable six-axis inertial measurement unit via recurrent neural networks. *Sensors* 21, 4 (2021), 1404.

[107] Virag Varga, Gergely Vakulya, Alanson Sample, and Thomas R. Gross. 2018. Enabling Interactive Infrastructure with Body Channel Communication. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 169 (jan 2018), 29 pages. doi:10.1145/3161180

[108] Anandghan Waghmare, Youssef Ben Taleb, Ishan Chatterjee, Arjun Narendra, and Shwetak Patel. 2023. Z-Ring: Single-Point Bio-Impedance Sensing for Gesture, Touch, Object and User Recognition. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 150, 18 pages. doi:10.1145/3544548.3581422

[109] Anandghan Waghmare, Qiuyue Xue, Dingtian Zhang, Yuhui Zhao, Shivan Mittal, Nivedita Arora, Ceara Byrne, Thad Starner, and Gregory D Abowd. 2020. UbiquiTouch: Self Sustaining Ubiquitous Touch Interfaces. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 27 (mar 2020), 22 pages. doi:10.1145/3380989

[110] Edward J. Wang, Tien-Jui Lee, Alex Mariakakis, Mayank Goel, Sidhant Gupta, and Shwetak N. Patel. 2015. MagnifiSense: inferring device interaction using wrist-worn passive magneto-inductive sensors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 15–26. doi:10.1145/2750858.2804271

[111] Jianqing Wang, Yuji Nishikawa, and Tsutayuki Shibata. 2009. Analysis of on-body transmission mechanism and characteristic based on an electromagnetic field approach. *IEEE transactions on microwave theory and techniques* 57, 10 (2009), 2464–2470.

[112] Wei Wang, Alex X. Liu, and Ke Sun. 2016. Device-Free Gesture Tracking Using Acoustic Signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking* (New York City, New York, USA) *(MobiCom '16)*. Association for Computing Machinery, New York, NY, USA, 82–94. doi:10.1145/2973750.2973764

[113] J.A. Ward, P. Lukowicz, G. Troster, and T.E. Starner. 2006. Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 10 (2006), 1553–1567. doi:10.1109/TPAMI.2006.197

[114] Erwin Wu, Ye Yuan, Hui-Shyong Yeo, Aaron Quigley, Hideki Koike, and Kris M. Kitani. 2020. Back-Hand-Pose: 3D Hand Pose Estimation for a Wrist-Worn Camera via Dorsum Deformation Network. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 1147–1160. doi:10.1145/3379337.3415897

[115] Wei Xia, Lin-Wei Ye, and Hong-Hai Liu. 2017. A gesture database of B-mode ultrasound-based human-machine interface. In *2017 International Conference on Machine Learning and Cybernetics (ICMLC)*, Vol. 1. IEEE, 118–122.

[116] Yang Xingchen, Dalin Zhou, Yu Zhou, Youjia Huang, and Honghai Liu. 2018. Towards Zero Re-Training for Long-Term Hand Gesture Recognition via Ultrasound Sensing. *IEEE Journal of Biomedical and Health Informatics* PP (08 2018), 1–1. doi:10.1109/JBHI.2018.2867539

[117] Chenhan Xu, Bing Zhou, Gurunandan Krishnan, and Shree Nayar. 2023. AO-Finger: Hands-Free Fine-Grained Finger Gesture Recognition via Acoustic-Optic Sensor Fusing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 306, 14 pages. doi:10.1145/3544548.3581264

[118] Jipeng Yan, Xingchen Yang, Xueli Sun, Zhenfeng Chen, and Honghai Liu. 2019. A Lightweight Ultrasound Probe for Wearable Human–Machine Interfaces. *IEEE Sensors Journal* 19, 14 (2019), 5895–5903. doi:10.1109/JSEN.2019.2905243

[119] Chouchang Jack Yang and Alanson P Sample. 2017. Em-comm: Touch-based communication via modulated electromagnetic emissions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–24.

[120] Jaeyoung Yang, Joonwhan Lee, and Joongmin Choi. 2011. Activity recognition based on RFID object usage for smart mobile devices. *Journal of Computer Science and Technology* 26, 2 (2011), 239–246.

[121] Xingchen Yang, Xueli Sun, Dalin Zhou, Yuefeng Li, and Honghai Liu. 2018. Towards Wearable A-Mode Ultrasound Sensing for Real-Time Finger Motion Recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26, 6 (2018), 1199–1208. doi:10.1109/TNSRE.2018.2829913

[122] Jia Zeng, Yu Zhou, Yicheng Yang, Jipeng Yan, and Honghai Liu. 2022. Fatigue-Sensitivity Comparison of sEMG and A-Mode Ultrasound based Hand Gesture Recognition. *IEEE Journal of Biomedical and Health Informatics* 26, 4 (2022), 1718–1725. doi:10.1109/JBHI.2021.3122277

[123] Ming Zeng, Le T Nguyen, Bo Yu, Ole J Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. 2014. Convolutional neural networks for human activity recognition using mobile sensors. In *6th international conference on mobile computing, applications and services*. IEEE, 197–205.

[124] Shibo Zhang, Yaxuan Li, Shen Zhang, Farzad Shahabi, Stephen Xia, Yu Deng, and Nabil Alshurafa. 2022. Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors* 22, 4 (2022), 1476.

[125] Shibo Zhang, Yuqi Zhao, Dzung Tri Nguyen, Runsheng Xu, Sougata Sen, Josiah Hester, and Nabil Alshurafa. 2020. Necksense: A multi-sensor necklace for detecting eating activities in free-living conditions. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 4, 2 (2020), 1–26.

[126] Yifan Zhang, Congqi Cao, Jian Cheng, and Hanqing Lu. 2018. EgoGesture: A New Dataset and Benchmark for Egocentric Hand Gesture Recognition. *IEEE Transactions on Multimedia* 20, 5 (2018), 1038–1050. doi:10.1109/TMM.2018.2808769

[127] Bing Zhou, Matias Aiskovich, and Sinem Guven. 2021. Acoustic Sensing-based Hand Gesture Detection for Wearable Device Interaction. arXiv:2112.05986 [cs.HC]

[128] Feifei Zhou, Xiangyu Li, and Zhihua Wang. 2020. Efficient High Cross-User Recognition Rate Ultrasonic Hand Gesture Recognition System. *IEEE Sensors Journal* 20, 22 (2020), 13501–13510. doi:10.1109/JSEN.2020.3004252

[129] Lu Zongxing, Cai Shaoxiong, Bingxing Chen, Zhoujie Liu, Lin Guo, and Ligang Yao. 2022. Wearable Real-Time Gesture Recognition Scheme Based on A-Mode Ultrasound. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* PP (09 2022), 1–1. doi:10.1109/TNSRE.2022.3205026