Grab-n-Go: On-the-Go Microgesture Recognition with Objects in Hand

CHI-JUNG LEE, Cornell University, USA
JIAXIN LI, Cornell University, USA
TIANHONG CATHERINE YU, Cornell University, USA
RUIDONG ZHANG, Cornell University, USA
VIPIN GUNDA, Cornell University, USA
FRANÇOIS GUIMBRETIÈRE, Cornell University, USA
CHENG ZHANG, Cornell University, USA

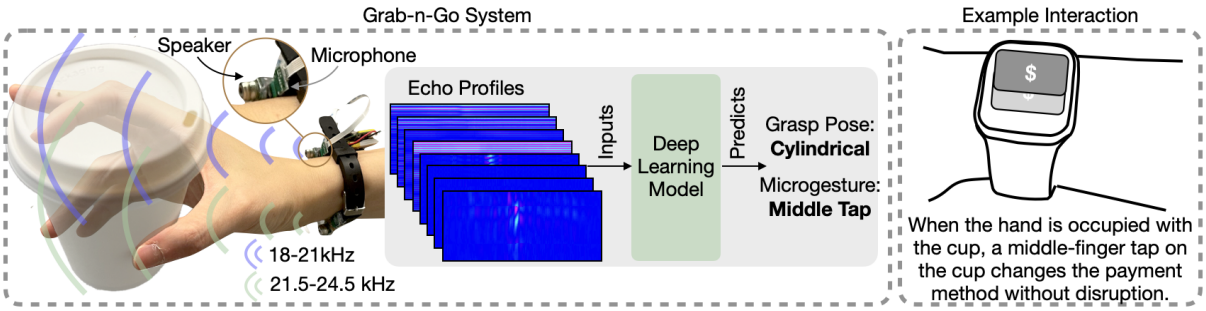

Fig. 1. We propose Grab-n-Go, a wristband that can recognize 30 microgestures across 35 various everyday objects. With two speaker-microphone pairs on each side of the wrist emitting and receiving the acoustic waves ranging 18-21 kHz and 21.5-24.5 kHz, respectively, echo profiles can be created to infer the microgestures using a customized deep learning model. Grab-n-Go enables gestural control when the hands are occupied.

As computing devices become increasingly integrated into daily life, there is a growing need for intuitive, always-available interaction methods — even when users' hands are occupied. In this paper, we introduce Grab-n-Go, the first wearable device that leverages active acoustic sensing to recognize subtle hand microgestures while holding various objects. Unlike prior systems that focus solely on free-hand gestures or basic hand-object activity recognition, Grab-n-Go simultaneously captures information about hand microgestures, grasping poses, and object geometries using a single wristband, enabling the recognition of fine-grained hand movements occurring within activities involving occupied hands. A deep learning framework processes these complex signals to identify 30 distinct microgestures, with 6 microgestures for each of the 5 grasping poses. In a user study with 10 participants and 25 everyday objects, Grab-n-Go achieved an average recognition accuracy of 92.0%. A follow-up study further validated Grab-n-Go's robustness against 10 more challenging, deformable objects. These results

Authors' Contact Information: Chi-Jung Lee, Cornell University, Ithaca, New York, USA, cl2358@cornell.edu; Jiaxin Li, Cornell University, Ithaca, New York, USA, jl2726@cornell.edu; Tianhong Catherine Yu, Cornell University, Ithaca, New York, USA, ty274@cornell.edu; Ruidong Zhang, Cornell University, Ithaca, New York, USA, rz379@cornell.edu; Vipin Gunda, Cornell University, Ithaca, New York, USA, vg245@cornell.edu; François Guimbretière, Cornell University, Ithaca, New York, USA, fvg3@cornell.edu; Cheng Zhang, Cornell University, Ithaca, New York, USA, chengzhang@cornell.edu.

ACM 2474-9567/2025/9-ART99
https://doi.org/10.1145/3749469

underscore the potential of Grab-n-Go to provide seamless, unobtrusive interactions without requiring modifications to existing objects. The complete dataset, comprising data from 18 participants performing 30 microgestures with 35 distinct objects, is publicly available at https://github.com/cjlisalee/Grab-n-Go_Data with the DOI: https://doi.org/10.7298/7kbd-vv75.

CCS Concepts: • **Human-centered computing → Interaction devices**.

Additional Key Words and Phrases: Wearable, Acoustic Sensing, Smartwatch, Gesture

ACM Reference Format:
Chi-Jung Lee, Jiaxin Li, Tianhong Catherine Yu, Ruidong Zhang, Vipin Gunda, François Guimbretière, and Cheng Zhang. 2025. Grab-n-Go: On-the-Go Microgesture Recognition with Objects in Hand. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 3, Article 99 (September 2025), 27 pages. https://doi.org/10.1145/3749469

1 INTRODUCTION

As computing devices become increasingly integrated into daily life, the need for always-available, unobtrusive interaction methods continues to grow. While hands are the primary means of interaction with these devices — through actions like typing, swiping, and gesturing — in everyday scenarios, hands are frequently occupied with holding, carrying, or using objects, making traditional input methods impractical. This presents a challenge, as the need to interact with computing devices persists, particularly for quick actions like answering phone calls or controlling smart home devices.

Subtle hand gestures, or microgestures, performed while holding objects offer a promising solution for always-available input [45, 47]. However, recognizing these microgestures in the presence of objects introduces significant challenges. One approach involves instrumenting objects with sensors [43, 70], but this can be costly and not feasible for everyday objects, especially consumables. Wearable-based methods can potentially overcome this limitation by eliminating the need for object modifications [13, 38, 39, 48]. Yet, existing wearable solutions typically focus solely on tracking hand movements, ignoring the objects themselves. This limitation reduces their generalizability across various objects. As a result, most prior work has been restricted to recognizing microgestures across a relatively small set of objects, typically around 10 or fewer.

Recent advancements in wearable computing have demonstrated the potential of active acoustic sensing for tracking fine-grained movements of different body parts [19–21, 26–28, 33, 50, 67, 68]. More specifically, while previous work has successfully captured fine-grained hand movements and some hand-object interactions [17, 22, 62], these systems primarily focus on free-hand gestures or the identification of broader object-related activities. Notably, the recognition of hand microgestures performed while holding objects remains an underexplored area within this domain.

The presence of objects in hand introduces unique challenges for active acoustic sensing:

(1) **Diverse Object Geometries**: Varying object shapes lead to different grasping poses and signal variations. The challenge is to build a system that can generalize microgesture recognition across a wide range of object geometries.
(2) **Occlusion**: Objects in hand can block sensor views, interfering with signal capture — especially when detecting subtle hand and finger movements. The challenge is to reliably capture and distinguish these subtle finger motions despite significant acoustic signal reflections caused by various object shapes and grasping poses.

Despite these challenges, we hypothesize that air-borne active acoustic sensing can simultaneously capture information about hand-grasping poses, object geometries, and microgestures using a single wristband. This approach offers a promising solution for robust microgesture recognition across a wide range of objects without requiring modifications to them.

This paper explores the following research question:

- How can we develop a wristband with active acoustic sensing that reliably recognizes a rich set of microgestures while a user holds various objects in different grasping poses?

To answer this question, we present Grab-n-Go, a wristband-based system designed to recognize 30 hand microgestures while holding various objects. Grab-n-Go employs active acoustic sensing by using two embedded speakers to emit inaudible acoustic waves toward the hand and the object. Two microphones on the wristband capture the reflected waves, which form unique patterns corresponding to each microgesture.

We evaluated Grab-n-Go in a user study with 10 participants. The study examined 5 distinct grasping poses, defined by Schlesinger's grasp taxonomy [47]. Each of these grasping poses involved 6 microgestures, resulting in a total of 30 microgestures (Fig. 2). To ensure generalizability, we included 5 diverse objects per grasping pose, totaling 25 objects. Each participant performed 6 microgestures 24 times for 2 randomly assigned objects within each grasping pose, resulting in 1,440 microgesture samples per participant. Our model achieved an average accuracy of 92.0% across all objects, demonstrating the effectiveness of Grab-n-Go in recognizing hand microgestures while hands are occupied.

To further explore system boundaries, we conducted a follow-up study with 8 additional participants specifically targeting 10 more challenging, deformable objects. These objects introduce greater variability across testing sessions and continuous shape transformations during microgesture execution. When evaluating on a combined dataset from both studies, and assessing performance on one object per grasping pose (compared to two in the initial study), we observed recognition accuracies of 95.0% for non-deformable objects and 92.9% for deformable objects. This suggests that Grab-n-Go benefits from a larger and more diverse training dataset and exhibits promising capability in handling the increased complexity introduced by deformable objects. The complete dataset with 18 participants and 35 objects is publicly available at https://github.com/cjlisalee/Grab-n-Go_Data with the DOI: https://doi.org/10.7298/7kbd-vv75.

In summary, Grab-n-Go is the first wearable sensing device specifically designed to recognize a rich set of microgestures while holding various objects, paving the way for more natural and always-available hand interactions. We conclude with a discussion of the challenges and design implications for future wearable systems.

The contributions of this paper are:

- We show that active acoustic sensing on a wristband can effectively recognize hand microgestures even when users are holding various objects.
- We conducted a user study with 10 participants performing 30 microgestures across 5 distinct grasping poses and 25 different objects. This was followed by another study with 8 participants using 10 deformable objects. The two studies validated the system's robustness and generalizability.
- We released the dataset, comprising a total of 20,160 microgesture instances performed by 18 participants across 35 distinct objects, to facilitate further research in this domain.
- We provide practical design implications for future wearable devices, outlining the challenges and opportunities for enabling seamless, always-available hand interactions even when hands are occupied.

2 RELATED WORK

Sensing hand microgestures while holding objects is challenging due to the subtle nature of these gestures and the wide variety of objects the hand may engage with. Prior research mainly focuses on two approaches, depending on sensor placement: instrumenting the objects or instrumenting the hands. Approaches that instrument the objects integrate sensors directly into the objects themselves, enabling them to detect microgestures, while approaches that instrument the hands rely on wearable sensors placed on the hand to capture microgestures. In this section, we explore each of these approaches in detail and discuss their relationship to our proposed Grab-n-Go system.

Table 1. Comparison with Prior Work

	Technique	Algorithm	Form Factor	Gestures	Objects	Performance	Remounting
Saponas et al. [39]	EMG	SVM	Armband	4 (Finger Press)	2	85% (tumbler) 88% (bag)	×
Rudolph et al. [38]	Capacitive Sensing	LDA	Wristband	6 (Object Interaction)	6	99%	×
VibAware [13]	Active & Passive Acoustic Sensing	SVM	Wristband + Ring	12	4 (3D-Printed Prop)	85.7%	×
SparseIMU [48]	Finger Joint IMUs	RF	On-Skin	19	12	0.93 F1 Score	✓
Grab-n-Go	**Active Acoustic Sensing**	**ResNet**	**Wristband**	**30**	25	92%	✓

2.1 Gesture Sensing Instrumenting the Objects

To enable microgesture sensing on objects, researchers have explored embedding sensors directly into various objects. One approach leverages the inherent conductivity of objects for touch sensing by connecting them to sensor boards [40, 70]. However, since many everyday objects are not naturally conductive, alternative methods have been developed, such as applying touch sensors to the surface [9, 36, 41, 56, 70] or embedding them inside the objects themselves [32, 51]. To automate the fabrication process, some researchers have focused on integrating touch sensors during manufacturing. Capricate [43] and MetaSense [4] enable touch sensing on 3D-printed objects by incorporating capacitive touch sensors during the 3D-printing process. While these approaches enable gestural user interfaces on various objects and support object-oriented interactions, they face a significant limitation: the need to instrument every object. This requirement presents a significant challenge for widespread adoption, as it is impractical to instrument all the objects people encounter in daily life.

To tackle the challenge, researchers have explored instrumenting more ubiquitous objects or materials. One example is cords, which are commonly found in everyday life — such as jacket drawstrings, charging cables, and bracelets. By integrating sensors into cords, these everyday objects can be enhanced with sensing capabilities [1, 15, 30, 31, 44]. In addition, textiles, which are prevalent in various aspects of daily life, from furniture to clothing, have been proposed as a key medium for ubiquitous computing. E-textiles, in particular, have shown significant potential for gesture recognition by utilizing fabric-based sensors to detect hand movements [58, 59, 61]. However, despite the potential of these innovations, the number of everyday objects people frequently use still far exceeds those that could be instrumented.

Instrumenting every object for microgesture sensing presents challenges. It is impractical for consumables and widely used everyday objects, as it would require constant modification and maintenance. Furthermore, augmenting objects with embedded sensors often needs significant customization, which can be complex, costly, and may not scale effectively across diverse objects. In contrast, instrumenting the hand with a wearable device eliminates the need to alter objects, offering a more flexible and ubiquitous solution for recognizing hand gestures while holding various objects. By focusing on the user rather than the object, wearable-based approaches enable seamless and consistent microgesture recognition across different scenarios. In the next subsection, we explore gesture-sensing approaches that instrument the hands.

2.2 Gesture Sensing Instrumenting the Hands

Given watches' long-standing social acceptance and minimal disruption to daily routines, wrist-worn devices, which can potentially be integrated into watches, or smartwatches, have emerged as a prime focus. In this subsection, we focus on wrist-worn devices for hand gesture recognition.

Recognizing hand gestures [2, 7, 10, 11, 14, 18, 29, 34, 37, 49, 52, 63, 65, 66, 69, 71] or even continuously tracking the hand poses [8, 12, 16, 17, 23, 24, 57, 60, 64] has been extensively studied. However, most of these works are confined to free-hand gestures and do not investigate how the system performs when hands are occupied by

objects. The presence of objects in the hand can introduce occlusion and noise, hindering gesture recognition. For instance, FingerTrak [8] experienced significant performance degradation in hand tracking when the hand is holding small objects, underlining the challenges posed by in-hand objects.

Some other methods have been proposed to recognize hand gestures with objects in hand. Leveraging forearm electromyography (EMG), Saponas et al. [39] developed an armband capable of distinguishing pinching movements when holding a tumbler or a handbag. Their approach utilized features extracted from filtered 1D EMG signals to train a support vector machine (SVM) classifier. Rudolph et al. [38] presented a wristband that can recognize force, grasp, and object manipulation based on capacitive sensing. This system extracted statistical features from 2D heatmaps representing the spatial distribution of capacitive magnitude signals and trained linear discriminant analysis (LDA) models. VibAware [13] supports tap and swipe when grasping different shapes of objects using bio-acoustic sensing. Their method extracted features from 1D filtered bio-acoustic signals to train an SVM classifier. SparseIMU [48] is a platform that supports the required IMUs on hand for sensing different gesture sets, employing features extracted from 1D filtered IMU signals to train a random forest (RF) classifier.

However, these sensing approaches focus mostly on the internal status of the hands yet lack information on the objects as well as the interaction between hands and objects, i.e., grasping poses. As a result, these works mostly investigate a small number of objects that have similar shapes or sizes. The ability of these systems to function across various objects has not been extensively investigated. In contrast, Grab-n-Go leverages active acoustic sensing, which captures information about not only the hands but also the objects around them. Therefore, Grab-n-Go can consistently work across various everyday objects. Notably, the rich spatiotemporal information contained in our 2D feature maps generated by our approach, the echo profiles, presents greater complexity than the signals processed in prior work, necessitating more sophisticated approaches beyond the generic machine learning methods employed in previous work. Our analysis demonstrates that advanced deep learning architectures more effectively extract and leverage the complex patterns inherent in these acoustic feature maps and generate better results.

The closest work to ours is EchoWrist [17], which employs active acoustic sensing on a wristband for tracking free-hand poses and recognizing basic hand-object interactions. However, as discussed previously, recognizing subtle hand gestures while holding objects presents unique challenges, including grasping variations and signal occlusions, which have not been addressed in prior work. While EchoWrist can identify broader hand-object activities. e.g., holding chopsticks or stirring with chopsticks, it does not capture the fine-grained hand movements occurring within these activities. To the best of our knowledge, Grab-n-Go is the first wearable device to demonstrate the use of active acoustic sensing for recognizing a rich set of hand microgestures while holding various objects, showing clear improvements over existing microgesture recognition approaches for occupied hands.

3 DESIGN CONSIDERATIONS

To address the research question: **"How can we develop a wristband with active acoustic sensing that reliably recognizes a rich set of microgestures while a user holds various objects in different grasping poses?"**, and ultimately enable always-available gestural input in daily life while minimizing disruption to ongoing activities, we propose the following design considerations:

3.1 Microgesture Design and Generalization across Various Objects

In everyday lives, people's hands are frequently occupied by a wide range of objects, which vary significantly in shape, material, size, and weight [46, 72]. This inherent diversity poses a significant challenge for consistently recognizing the same microgesture across different objects. Generalizing microgesture recognition to work effectively across such a wide variety of objects remains a key research problem.

Fig. 2. The object and microgesture sets used for evaluating the capability of Grab-n-Go.

Prior research [25, 42, 47] suggests that microgestures are largely determined by the grasping poses and the geometry of the objects being held. This indicates that even when objects differ in shape, material, size, or weight, they may still allow for the same set of microgestures if they are held using similar grasping poses. In other words, the way human hands hold an object plays a more critical role in determining microgestural possibilities than the object's specific characteristics. By prioritizing grasping poses rather than individual object properties, the system can generalize microgesture recognition across a wide range of objects. This approach significantly reduces the need for extensive object-specific training data, enhancing scalability and adaptability. Instead of training the system on a vast number of objects, it can focus on a finite set of grasping poses that naturally emerge across different interactions, making microgesture recognition more efficient and robust in diverse real-world scenarios.

To ensure generalizability across various objects, Grab-n-Go aims to classify the same set of microgestures when performed with the same grasping pose regardless of the specific object being held. This approach assumes that for all objects typically grasped in a given pose, the available microgestures remain consistent. By focusing on grasping poses rather than object-specific characteristics, Grab-n-Go can reduce the need for extensive

object-specific training data and enhance adaptability across a diverse range of everyday objects. Following prior research [3, 13, 39, 48], we adopt the grasping poses classification system defined by Schlesinger [42] as the basis for grouping microgestures. However, we exclude the "lateral" grasp, as Sharma et al. [47] point out that this grasping pose constrains the most dexterous fingers — the thumb and pointer finger — making it challenging to perform microgestures with the remaining fingers. By focusing on grasping poses that allow for finer finger movements, we optimize microgesture sets while maintaining broad applicability across different everyday scenarios.

Our microgesture set design (Fig. 2) is informed by prior research on microgestures performed with occupied hands [13, 38, 47]. We define a set of 5 dynamic microgestures for each grasping pose. These microgestures are carefully selected based on their feasibility, distinctiveness, and potential for seamless execution while holding various objects. In addition to the dynamic microgestures, we incorporate a static holding state, in which the user simply holds an object without other movements. This state serves as a neutral baseline, preventing unintended activations and enabling practical, real-world applications where microgesture input should only be recognized when explicitly performed. We call it the Hold microgesture in the rest of the paper. In total, our microgesture set consists of 30 unique microgestures, covering a diverse range of grasping poses and objects. The 5 grasping poses and their 6 corresponding microgestures are as follows:

- Cylindrical:
  - Grasping Pose: An open fist grip for cylindrical objects (*e.g.*, paper cups) where the thumb is positioned on one side of the object and the other four fingers on the opposite side.
  - Constraints: The thumb and at least two fingers are required to maintain the grip, with the pointer and middle fingers having greater dexterity. Therefore, the proposed microgestures involve movements of the pointer finger, middle finger, and wrist.
  - Proposed Microgestures: Hold, Pointer In, Pointer Tap, Middle Tap, Wrist Right, and Wrist Left.
- Spherical:
  - Grasping Pose: An open fist grip for spherical objects (*e.g.*, tennis balls), where all the fingers are evenly distributed around the objects.
  - Constraints: The thumb and two additional fingers are essential for maintaining the hold. To avoid altering the grasping position while performing microgestures, movements primarily involve the pointer finger, middle finger, and wrist.
  - Proposed Microgestures: Hold, Pointer In, Pointer Tap, Middle Tap, Two-Finger Tap, and Wrist Up.
- Palmar:
  - Grasping Pose: A posture for holding flat, thick objects (*e.g.*, moving boxes), where the thumb stabilizes the object from the side while the other four fingers support it from underneath.
  - Constraints: To avoid dropping the object and consider finger dexterity, we propose microgestures using the thumb, pointer finger, and wrist.
  - Proposed Microgestures: Hold, Thumb Tap, Thumb In, Thumb Down, Pointer In, and Wrist Tap.
- Tip:
  - Grasping Pose: A grip for sharp and small objects (*e.g.*, pens), requiring at least two fingers for secure handling.
  - Constraints: The ring and pinky fingers have the most freedom of movement, though the ring finger is difficult to move independently. As a result, the proposed microgestures focus on the thumb, pinky finger, and wrist.
  - Proposed Microgestures: Hold, Thumb Tap, Thumb Right, Pinky Out, Wrist Left, and Wrist Right.
- Hook:

- Grasping Pose: A posture for carrying heavy objects with handles (*e.g.*, suitcase handles) where all fingers except the thumb are used to hook and secure the object.
- Constraints: The grip provides the most flexibility for thumb movements.
- Proposed Microgestures: Hold, Thumb Tap, Thumb Left, Thumb In, Pointer In, and Rotate.

To create a seamless and intuitive user experience, we designed all microgestures to start from the natural holding position and return to this initial state upon completion. For example, if the user is holding a tumbler and wants to perform the Pointer In microgesture, the user simply slides their pointer finger inward and then returns it to its original position. There is no need to first reposition the finger before executing the gesture, reducing cognitive effort and making interactions more fluid and natural.

3.2 The Choice of Sensing Technique and Form Factor

Hands are frequently occupied with objects during daily activities, presenting a significant challenge for wearable sensing systems. The presence of objects can obstruct sensors, leading to signal occlusion or undesired interference. To address this challenge, a sensing technique is required that not only mitigates the effects of occlusion but also leverages the unique signal characteristics introduced by object interference to enhance recognition. Ideally, the sensing technique should capture comprehensive information about grasping poses, hand microgestures, and object geometry simultaneously.

In addition, for microgesture recognition to be practical in everyday settings, factors such as cost, power efficiency, and user comfort must be carefully considered. An ideal sensing solution should adopt a widely accepted and minimally obtrusive form factor while integrating affordable, readily available sensors with low power consumption. By balancing accuracy, practicality, and usability, the system can support seamless and unobtrusive interaction in real-world scenarios.

Considering these factors, we propose Grab-n-Go, a wristband embedded with active acoustic sensing. Wristbands, which can be seamlessly integrated into smartwatches, have long been one of the most popular and widely accepted wearable form factors, offering a comfortable and unobtrusive user experience. More importantly, active acoustic sensing employs compact, low-cost sensors and has demonstrated promising results in estimating hand poses when hands are empty [17, 62]. With airborne acoustic signals, both the hand and the in-hand objects reflect the emitted waves, encoding information about their geometry. When a user performs microgestures, the resulting signal changes form distinct patterns that are independent of the specific object being held. This enables Grab-n-Go to capture a rich set of information, including object shape, grasping poses, and microgestures, all based on a single sensing modality. To the best of our knowledge, no existing system has explored this approach or achieved such comprehensive sensing capabilities. By jointly learning grasping poses, object properties, and microgestures, Grab-n-Go has the potential to recognize a wide range of hand microgestures across various objects — using only a single wristband.

4 THE DESIGN AND IMPLEMENTATION OF GRAB-N-GO

To recognize the microgestures while hands are occupied with various objects, we designed a compact wristband powered by active acoustic sensing and customized machine-learning inference pipelines. In this section, we provide a detailed overview of Grab-n-Go's hardware design, sensing principle, and machine learning pipeline.

4.1 Hardware Design

We designed Grab-n-Go to be a small, compact, and low-power device, ensuring its suitability for everyday use. The device is built into a silicone wristband, which is commonly used for watches, offering comfort and flexibility. To accommodate varying wrist sizes, the prototype is designed with an adjustable sensor layout, allowing for easy customization to fit different users.

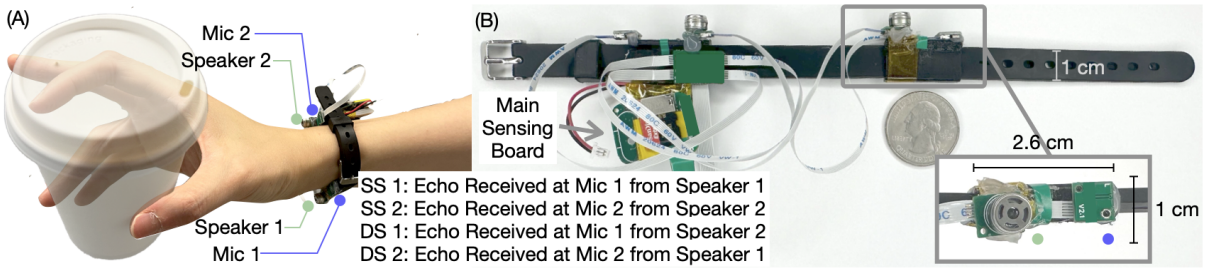

Fig. 3. Grab-n-Go Prototype.

The system incorporates two speaker-microphone pairs (OWR-06944T-16B and ICS-43434), which are mounted on customized printed circuit boards (PCBs) specifically designed for the sensors. These sensors are strategically positioned to face the hand, enabling them to capture detailed information about both the hand's movements and the object being held. This arrangement allows the system to detect subtle changes in the acoustic signals reflected by the hand and the object, which are crucial for accurate microgesture recognition. Each speaker-microphone pair is housed in a small 3D-printed case that securely connects the PCBs to the silicone wristband. This case is designed to slide along the wristband, enabling straightforward adjustment of the sensor placement to optimize performance for different users. The sensors are connected to a customized microcontroller module, which includes an SGW1110 module and an MAX98357A audio amplifier. These components are linked via flexible printed circuit (FPC) ribbons.

Powered by a LiPo battery, the system operates by having the microcontroller drive the speakers to emit sound waves while simultaneously collecting the reflected acoustic signals with the integrated microphones. Collected signal data can be stored on a microSD card for offline analysis or transmitted to a smartphone in real time via Bluetooth Low Energy (BLE) for immediate processing.

The two speakers emit acoustic signals of different frequency ranges: 18-21 kHz for one speaker and 21.5-24.5 kHz for the other. This frequency separation ensures that the signals captured by the microphones can be distinctly identified based on their respective frequencies. Specifically, the signals captured by the microphone placed on the Same Side (SS) and Different Side (DS) of the speaker can be differentiated by applying different band-pass filters (Fig. 3). Leveraging these four distinct acoustic signal travel paths — each corresponding to a different route the sound waves take from the speakers to the microphones — the system can capture comprehensive information about both the hand and the objects held in it (Fig. 4). This multi-path signal processing enables a richer, more nuanced understanding of the user's actions.

4.2 Theory of Operation

Grab-n-Go is powered by active acoustic sensing. As detailed in the Hardware Design Section (Sec. 4.1), we placed two speaker-microphone pairs on the wrist, facing the hand. These speakers emit frequency-modulated continuous waves (FMCW) towards the hand and the object being held. The geometry of both the hand-grasping poses and the objects creates a special reflection medium for the acoustic waves, resulting in distinct patterns in the captured signals. As shown in Fig. 4, when the user Holds different objects, the captured signals present different characteristics. However, due to the hand's closer proximity to the sensors, the grasping pose introduces a more dominant influence on the captured signal compared to the object itself. Notably, our grasping pose categorization method is based on the overall object geometry, ensuring that objects within the same grasp category tend to produce similar acoustic reflection patterns. This inherent characteristic of the system facilitates the generalization of microgesture recognition across a wide range of objects.

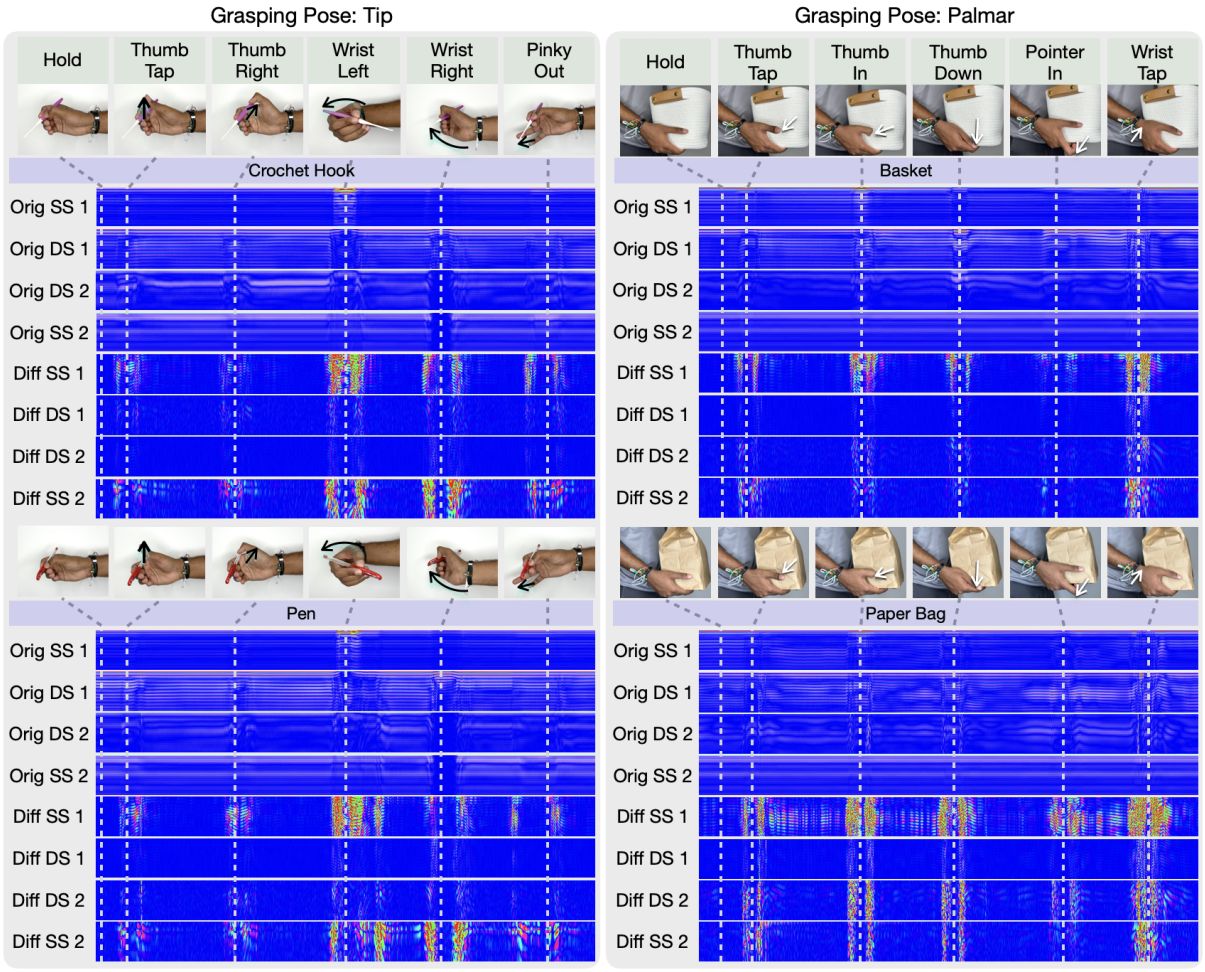

Fig. 4. Grab-n-Go Example Signals. The microphones capture the acoustic signal emitted from the speaker on the Same Side (SS) and Different Side (DS), resulting in 4 channels (= 2 (microphones) × 2 (speakers)) of Original (Orig) echo profiles. Subtracting the previous echo profile from the current one creates Differential (Diff) echo profiles, which focus on the movements. In total, 4 channels of original echo profiles and 4 channels of differential echo profiles are used for the system.

Unlike conventional passive audio analysis, the acoustic signal reflection patterns are formalized using a correlation-based FMCW (C-FMCW) approach called *echo profile analysis*, which is based on the cross-correlation between the transmitted and received acoustic signals [55]. This technique has proven effective in tracking body part movements when deployed on various wearable devices [17, 21, 27]. In Grab-n-Go, each frequency sweep has a duration of 12 ms. We perform cross-correlation between the transmitted signal and the band-pass filtered received signal to extract the signal strengths at different return times. Subsequently, by mapping these time-domain results into the distance domain, using the known speed of sound, we generate the *echo profiles* (Fig. 4).

In the echo profiles, each pixel's value represents the correlation strength, which reflects the intensity of the returned acoustic signal. The x-axis of the echo profile corresponds to time, with 12 ms per pixel, while the y-axis represents distance, which is 3.43 mm per pixel. A bright strip in the echo profile indicates a strong reflection at a specific distance. As observed in the example signals (Fig. 4), the echo profiles of larger objects, such as the Basket and Paper Bag, present broader and thicker bright strips compared to those of smaller objects like the Crochet Hook and Pen, reflecting the increased surface area for acoustic signal reflections. This approach differs from passive acoustic sensing techniques that typically employ Mel spectrograms to model human auditory perception by reweighting frequency bands according to psychoacoustic principles. Such representations would be suboptimal for our application as they compress precisely the high-frequency information critical for discriminating subtle finger movements. Instead, our echo profiles preserve the spatiotemporal reflection characteristics that directly correspond to physical microgesture execution, capturing the complex interplay between hand configuration and object geometry rather than ambient acoustic events.

Patterns within the echo profiles reveal changes in the distribution of reflection strengths over different distances and times. To isolate the movements of the hand during microgestures from constant environmental reflections and the static presence of held objects, we calculate *differential echo profiles*. This is achieved by subtracting the preceding pixel value from the current pixel in the echo profile, thereby emphasizing changes over time. As demonstrated in the example signals (Fig. 4), the differential echo profiles amplify the changes in echo profiles, which directly correspond to hand movements. For example, when the user is statically Holding the objects, the differential echo profiles present minimal patterns, whereas the patterns are obvious during the execution of dynamic microgestures. Notably, while different hand microgestures and grasping poses yield unique echo profile patterns, objects held in the same grasp pose tend to produce similar features. This characteristic is key to reliably recognizing hand microgestures across a variety of objects.

4.3 Machine Learning Pipeline

4.3.1 Input. To recognize microgestures, we developed a customized deep-learning pipeline. After the echo profile analysis, these microgestures are represented as different patterns in the *echo profiles*, which lay out as 2D feature maps, as demonstrated in Fig. 4. By cropping the echo profiles to the window of interest, we obtain input data with a size of $155 \times 70 \times 8$. This input tensor comprises 1.8 seconds of temporal data (155 pixels along the time axis), a 24 cm range of interest (70 pixels along the distance axis), and 8 stacked channels: four echo profiles and their corresponding four differential echo profiles.

4.3.2 Model Architecture. We propose an Encoder-Decoder model architecture. Given the proven effectiveness of Convolutional Neural Networks (CNNs) in decoding 2D information such as images, we select ResNet-18 [5] as the encoder backbone of our model. We incorporate an adaptive 2D average pooling layer with an output size of [1, 1], a dropout layer with a rate of 0.6 to prevent overfitting, and a fully connected layer with an output dimension of 30 to classify the 30 microgestures. Cross-entropy (CE) loss is employed as the optimization objective. The model is configured with an initial learning rate of 0.0002 and a batch size of 8.

4.3.3 Data Augmentation. To address potential variations in hand sizes and device positioning — including user differences and changes after remounting the device — we incorporate data augmentation techniques into our training process: (a) Vertical Shifting: Echo profiles were randomly shifted vertically by up to 6 pixels to account for slight variations in sensor-to-hand distances. (b) Amplitude Jitter: In 80% of training iterations, each pixel's intensity value was multiplied by a random factor between 0.95 and 1.05. This amplitude jitter introduces variability in the training data, preventing the model from overfitting to specific signal amplitudes and improving its robustness to noise.

5 USER STUDY

To evaluate Grab-n-Go's microgesture recognition performance when holding various objects (Sec. 3.1), we conducted a user study approved by the Institutional Review Board (IRB). We recruited 10 participants (3 self-identified as male, 7 as female; age: mean = 24.1, std = 4.04) with a wide variety of hand sizes and shapes (fingertip-to-wrist length: thumb: mean = 126 mm, std = 12 mm; pointer: mean = 170 mm, std = 12 mm; middle: mean = 177 mm, std = 11 mm; ring: mean = 165 mm, std = 12 mm; pinky: mean = 146 mm, std = 10 mm). Note that due to hardware issues, data from 3 of the original 13 participants was broken, resulting in their removal from the study and leading to 10 valid participants. Among the participants, one self-identified as ambidextrous, while the remaining nine were right-handed. Since the microgestures are designed to be easily performed with either hand, all participants were instructed to wear the device on their right wrist to maintain consistency in evaluation. Each study lasted approximately 2 hours, and participants were compensated US$25 for their time. The study followed a structured process: participants first completed a demographic survey and hand-size measurements, followed by the primary data collection (Sec. 5.2), and ended with a wearability survey to assess user comfort and experience.

5.1 Object Set

To evaluate Grab-n-Go's generalizability across a diverse range of objects, we selected five everyday objects for each grasping pose, as shown in Fig. 2. These objects exhibit a wide range of shapes, materials, sizes, and weights, reflecting the diversity encountered in real-world scenarios. To balance study duration while maximizing object diversity, each participant was randomly assigned 2 out of the 5 objects within each grasping pose category. Each object was then tested by four different participants, ensuring balanced data collection. Moreover, we carefully avoided repeated combinations of objects across participants to maintain objectivity and minimize potential biases in the evaluation.

5.2 Data Collection Procedure

5.2.1 Apparatus. The participants stood in front of a laptop (Apple MacBook Pro 14-inch, 2021) placed on a standing desk in a quiet study room. The laptop served multiple functions: it recorded hand movements using its built-in camera for ground truth label verification, displayed visual stimuli to signal the start of each data collection session, and provided on-screen instructions to guide participants through the process. All objects used for grasping were placed either on the desk or on the nearby ground, depending on their size, ensuring easy access while maintaining a natural interaction environment.

5.2.2 Data Collection Sessions. For each grasping pose in the order of Cylindrical, Hook, Tip, Palmar, and Spherical, each participant completed 1 practice session followed by 6 data collection sessions. To synchronize Grab-n-Go data with the laptop-recorded ground truth, the researcher initiated and concluded each session with a distinctive cue that is both audible and visible (a tumbler tap). During each session, the participant first performed 4 repetitions (1 repetition for practice sessions) of the 6 microgestures in a randomized order, using one of the assigned objects within the current grasping pose category. Each microgesture was performed in a 2-second window. Then, this process was repeated with the second assigned object. Between each session, the participant removed and re-wore the device under the researcher's guidance. In total, there were 2 (objects) × 5 (grasping poses) × 6 (sessions) × 6 (microgestures) × 4 (repetitions) = 1440 microgesture instances collected from each participant, resulting in a total of 14,400 microgesture instances across all participants. Following data review, 21 instances were relabeled and 45 instances were excluded due to incorrect microgesture execution.

5.3 Training Scheme

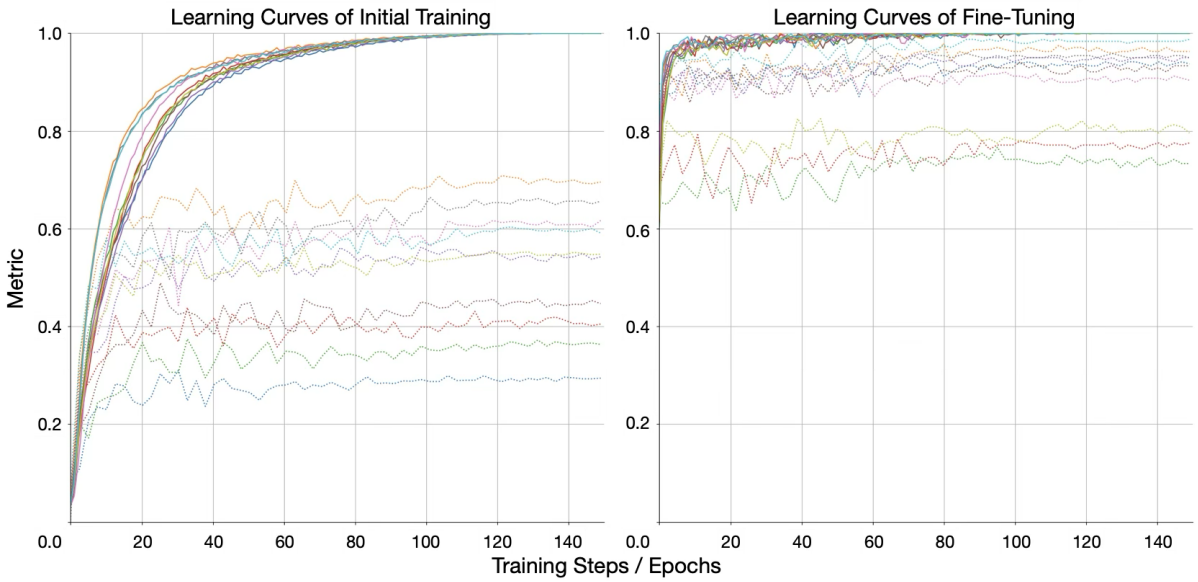

Fig. 5. Learning curves for the models trained on user study data. Learning curves for models trained on user study data. Each color represents a distinct participant's model, with solid lines indicating training performance and dotted lines showing testing performance.

5.3.1 Two-Step Training Scheme. To minimize training efforts for new users and enhance system performance, we implement a two-step training scheme. Although the system remains user-dependent, we optimize efficiency by fine-tuning a pre-trained user-independent model rather than training a customized model from scratch for each user.

Specifically, for each participant in the user study, we first trained a leave-one-participant-out model. This model was trained on data from all other participants, excluding the target participant, and the model was tested on the data from the excluded target participant. This phase lasts 150 epochs and produces a user-independent model that captures generalizable features of microgesture execution. In the second step, we fine-tuned this user-independent model on the target participant's data for an additional 150 epochs, adapting the model to the specific characteristics of the individual user. The choice of 150 epochs for both the initial training and fine-tuning phases was determined through empirical validation during our pilot study, where we observed that model performance typically plateaued with minimal fluctuation around this value for both training and testing sets (Fig. 5). While we did monitor validation performance, we did not formally implement early stopping based on a specific performance increase threshold.

5.3.2 Wearing Session Independence. While not entirely user-independent, Grab-n-Go supports wearing session independence. This crucial feature eliminates the need for repeated data collection and model retraining each time the user removes and re-wears the device. This is particularly important in real-world scenarios where users will inevitably remove the device for charging or other reasons. With wearing session independence, users only need to provide a one-time data collection upon initial device acquisition, mirroring the familiar process of finger ID or face ID registration.

To evaluate wearing session independence, each participant in the user study collected data across six sessions for each grasping pose. Between each session, the participant removed and re-wore the device, simulating

real-world usage scenarios. During the training phase, the model was trained on data from five of the six sessions and subsequently tested with the remaining session. To mitigate the potential influence of user familiarity with the microgestures over time, the final performance metric was calculated by averaging the results obtained from all possible combinations of training and testing sessions.

5.4 Results

5.4.1 Training Scheme. To ensure the study remained manageable for the participants, we limited the number of microgesture instances collected per participant to 24 per microgesture per object (= 4 (repetitions) × 6 (sessions)), with the device being remounted between sessions. Given the relatively small size of this dataset, we employed a two-step fine-tuning training scheme (Sec. 5.3.1) to assess the potential benefits of larger base datasets. Initially, a leave-one-participant-out (LOPO) model was trained for each participant using data from the remaining 9 participants. Subsequently, to evaluate the system's wearing session independence, this LOPO model was fine-tuned using the leave-one-session-out (LOSO) method (Sec. 5.3.2) for each participant. In each iteration, the LOPO model was fine-tuned using data from 5 sessions and tested on the held-out session for each grasping pose. This process was repeated 6 times, with each session serving as the held-out set in turn.

5.4.2 Microgesture Recognition Results. Overall, by averaging the results across all the participants and wearing sessions, the fine-tuning training process achieved an average accuracy of 92.0% in recognizing 30 microgestures performed on 25 different objects (Fig. 6). Importantly, each participant interacted with a unique and randomly assigned combination of objects for each grasping pose, ensuring no two participants encountered the exact same object set. In addition, our object set included a wide variety of shapes, materials, sizes, and weights, reflecting the complexity of real-world interactions. The system's ability to maintain high accuracy across 25 different objects highlights its strong cross-object generalizability.

Beyond the encouraging average accuracy of our system, the results also demonstrate a low average false-positive rate of 0.2%. This metric, calculated as the ratio of False Positives to the sum of False Positives and True Negatives, is particularly critical for the practical usability of microgesture-based interaction systems. False activations represent one of the most substantial barriers to user acceptance and system reliability in real-world contexts. A low false-positive rate signifies that the likelihood of the system erroneously detecting a microgesture when none was intended is minimal. This is important for reducing user frustration and fostering a reliable user experience. The low false-positive rate further underscores the feasibility of Grab-n-Go for real-world deployment, as it suggests a minimal tendency to trigger unintended actions, even when interacting with a diverse array of everyday objects. This reliability is essential for users to integrate microgestural input into their daily routines without concerns about spurious activations. However, we also want to admit that this low false-positive rate was achieved in a relatively controlled lab environment. Further experiments and studies will be needed to test the system in real-world scenarios where the false-positive rate can likely be higher due to the huge variance of daily body postures and noise in real-world settings.

5.4.3 Fine-tuning Results. To assess the impact of our two-step fine-tuning training scheme, we evaluated the system's performance by directly training user-dependent models using only the data from each of the 10 individual participants. This purely within-user training involved using 5 sessions of data for training and the remaining 1 session for testing, all belonging to the same participant. Averaging the results across all participants and wearing sessions, this direct training approach yielded an average accuracy of 86.19%, a 5.81% decrease compared to the fine-tuning results (92.0%). This difference underscores the benefit of leveraging the foundation model trained on a broader dataset. Furthermore, it suggests that with the acquisition of even larger and more diverse datasets in the future, there is considerable potential for further performance improvements.

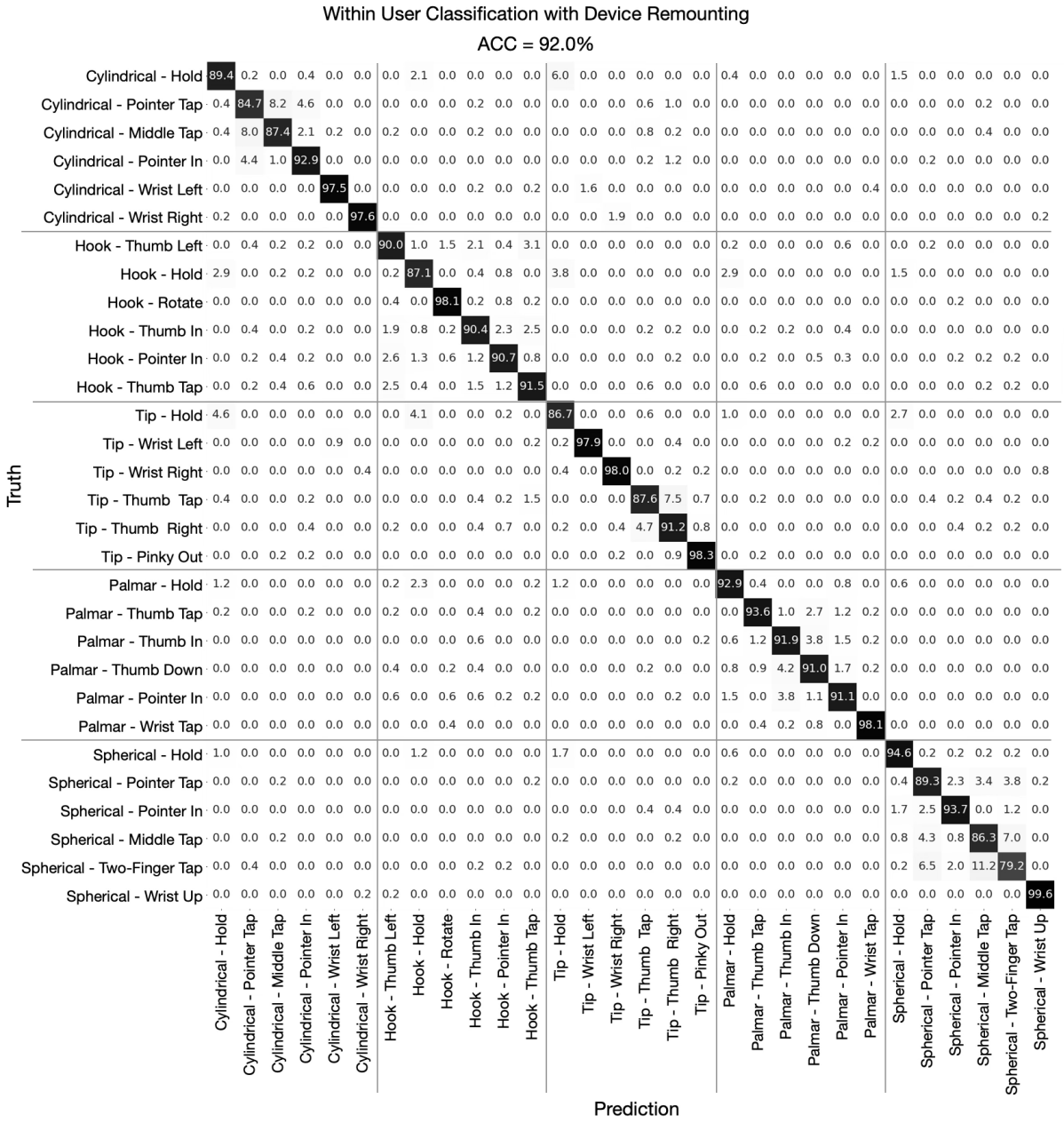

Fig. 6. The confusion matrix of the study results.

Since collecting personalized training data can be time-consuming and inconvenient, reducing the amount of required data is crucial for improving user adoption. To minimize the amount of training data required from new users, we investigated how much data is necessary to fine-tune the model without significantly compromising

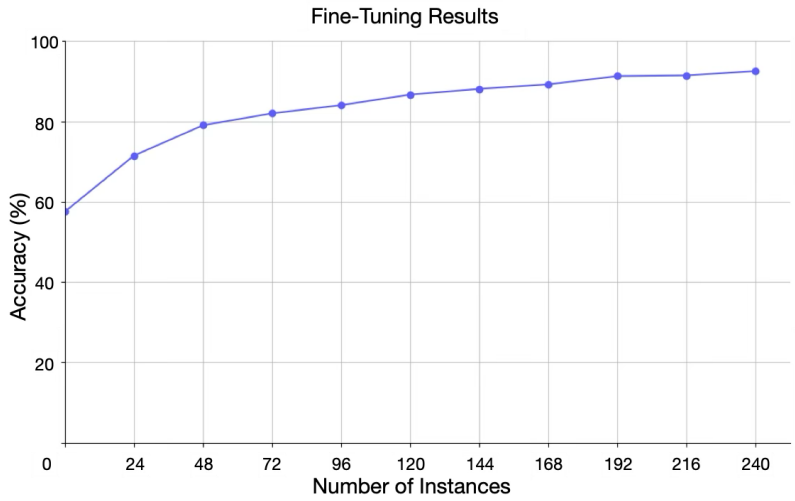

Fig. 7. The performance when fine-tuning the model with different amounts of data.

performance. We systematically varied the number of sessions, or microgesture instances, used during the fine-tuning process to assess its impact on recognition accuracy.

As shown in Fig 7, with a fully user-independent model, where no personalized data from the new user was used, the average accuracy was 57.5%, indicating that while the model captured some generalizable features, it struggled to adapt to individual differences. However, when incorporating just one session of data (24 microgesture instances per grasping pose), average accuracy increased to 71.4%, demonstrating that even a small amount of user-specific data enhances performance. The model exceeded 80% accuracy with three sessions (72 instances) and surpassed 90% accuracy with eight sessions (192 instances).

These results suggest a promising potential to reduce the required fine-tuning data. We believe that as the base user-independent model is trained with more data, the amount of data needed from new users can be further reduced, making Grab-n-Go more practical for real-world deployment.

5.4.4 Object-independent Results. To further investigate the generalizability of Grab-n-Go across various objects, we conducted an object-independent evaluation. Within each grasping pose category, we trained the model using data from four objects and tested it on the remaining object. This process was repeated for all the objects.

The average accuracy across the 25 objects is 85.3% (Cylindrical: 85.1%, Hook: 81.0%, Tip: 93.7%, Palmar: 87.1%, and Spherical: 79.5%). It is important to note that each of these object-independent models was trained on a relatively limited dataset of 576 microgesture instances (= 6 (microgestures) × 4 (repetitions) × 6 (sessions) × 4 (participants)) are used for training. In addition, the data came from different participants, introducing more variability and potential uncertainty, which could affect the performance. Given these factors, we believe that further improvements could be achieved by incorporating more data from the same participant or by including additional objects within the same grasping pose. This would help the model learn a richer set of features, ultimately boosting its performance and robustness across various objects and users.

5.4.5 Discussion. Despite the use of similar microgestures for different grasping poses, e.g., Pointer Tap was used for Cylindrical and Spherical, and Pointer In was used for Cylindrical, Spherical, Palmar, and Hook, Grab-n-Go successfully differentiated between them. This demonstrates the system's ability to recognize not only hand movements but also grasping poses. Analyzing the confusion matrix, incorrect predictions primarily occurred

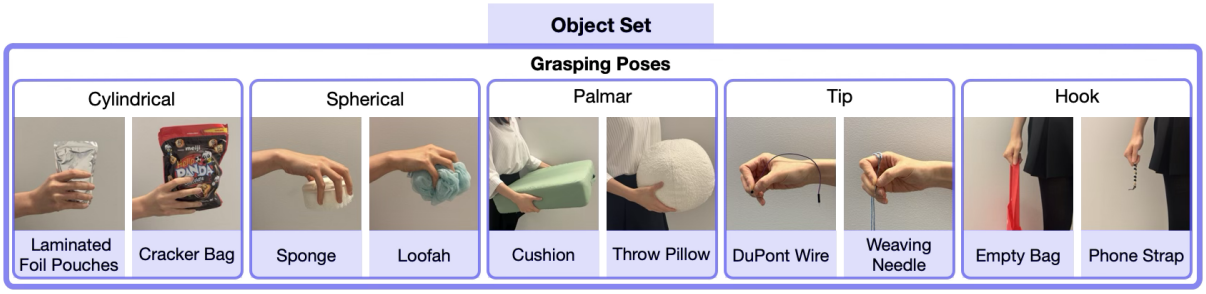

Fig. 8. The object set of the follow-up study.

within the same grasping pose category. This indicates that Grab-n-Go effectively classified distinct grasping poses, suggesting potential for further refinement by fine-tuning the model on specific grasping pose categories to enhance performance.

Our findings indicate that wrist-related microgestures exhibit the highest recognition accuracy. When examining the echo profiles (Fig. 4), these microgestures produce the most distinguishable signals, likely due to their proximity to the sensors and relatively large scale of movements on the palm. Despite being farther from the sensors and generating weaker signals, finger-related microgestures still achieve satisfactory performance. This suggests the potential for incorporating additional finger-based microgestures in future iterations.

Pointer Tap and Middle Tap exhibited the highest levels of confusion, which is unsurprising given the proximity of these two fingers and the similarity of their movements. In addition, between each session, the participant not only removed and re-wore the device but also re-held the object, introducing additional sources of variation, which was designed to simulate real-world scenarios. Despite these challenges, the majority of recognition accuracies remained above 92%. When analyzing performance within each grasping pose, Tip exhibited the least confusion, while Spherical demonstrated the most. This observation aligns with the observation that microgestures performed closer to the sensors can yield better accuracy. This insight provides valuable guidance for designing future wrist-worn active acoustic sensing devices and microgesture sets.

Overall, according to our user study evaluation results, Grab-n-Go effectively recognizes 30 microgestures across 25 distinct everyday objects using only a single wristband, positively supporting our proposed research question as described earlier in the paper.

5.5 Follow-up Study with Deformable Objects

While Grab-n-Go effectively captures object and hand shape for microgesture recognition, deformable objects present a challenge to sensing accuracy. Unlike rigid objects that maintain consistent acoustic reflection patterns, deformable materials introduce variability through shape distortions that occur during natural manipulation. Consequently, we conducted a follow-up study to assess the impact of these shape variations on performance.

This study involved 8 right-handed participants (3 self-identified as male, 5 as female; age: mean = 24.25, std = 3.45; fingertip-to-wrist length: thumb: mean = 131 mm, std = 8 mm; pointer: mean = 171 mm, std = 7 mm; middle: mean = 179 mm, std = 9 mm; ring: mean = 167 mm, std = 8 mm; pinky: mean = 144 mm, std = 9 mm). Consistent with our initial protocol, participants wore the prototype device on their right wrist throughout the study. Each study lasted approximately 1.5 hours, and participants received compensation of US$25 for their participation.

5.5.1 Study Setup. Maintaining the same process and apparatus as the initial study, we introduced a different object set comprising two deformable everyday items for each grasping pose (Fig. 8). These objects were specifically chosen due to two key challenges: (1) their shape varied with each grasp, and (2) they deformed further during

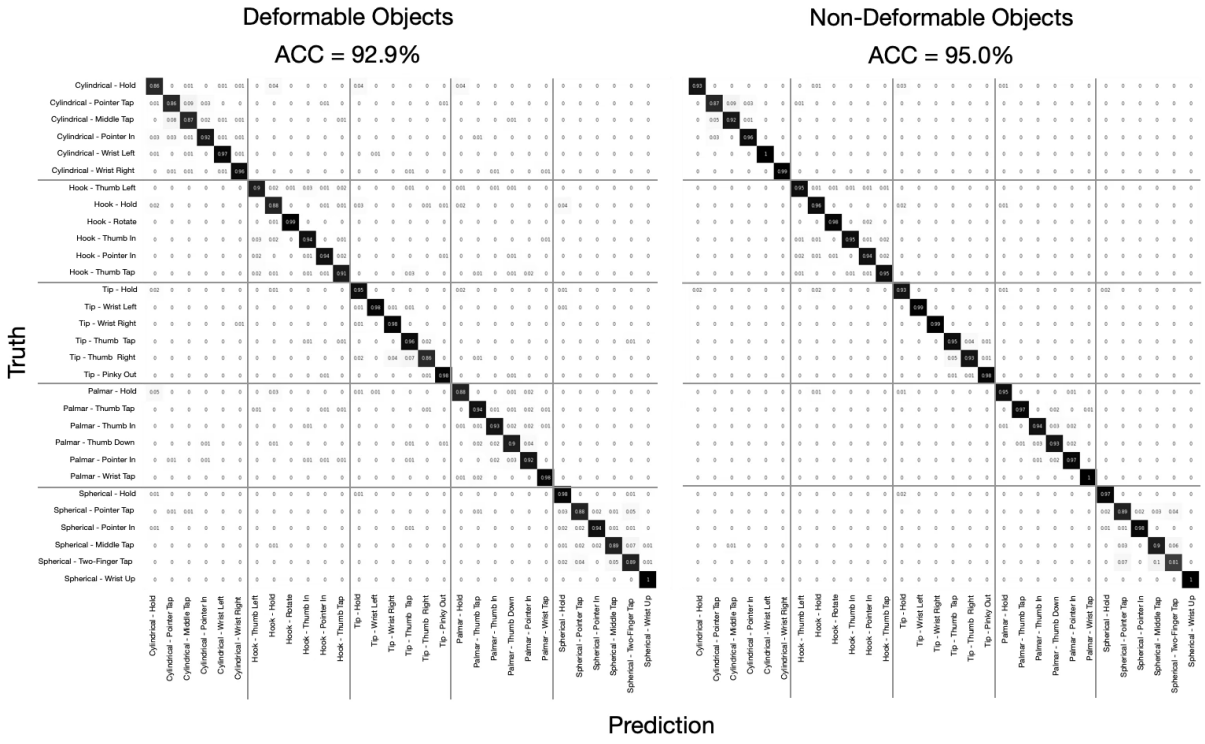

Fig. 9. The confusion matrix of the study results of the follow-up study.

microgesture execution. Each participant interacted with one object per grasping pose category, and each object was evaluated by four distinct participants, aligning with the initial study's protocol.

5.5.2 Data Collection Procedure. This study mirrored the procedure of the initial study, with the key modification that participants used a single object for all sessions within a specific grasping pose category. Consequently, each of the eight participants contributed 720 microgesture instances (= 1 (objects) × 5 (grasping poses) × 6 (sessions) × 6 (microgestures) × 4 (repetitions)), resulting in a dataset of 5,760 microgesture instances across all eight participants. Following data review, 30 instances were relabeled and 38 instances were excluded due to incorrect microgesture execution.

5.5.3 Results. Using the data from the initial study and the follow-up study, we created a joint dataset incorporating data from all 18 participants. Employing our two-step fine-tuning training scheme (Sec. 5.3.1), we initially trained 18 leave-one-participant-out (LOPO) models on this combined dataset. Note that as the follow-up study tested only a single object per grasping pose per participant, the fine-tuning phase for participants from the initial study utilized only the data corresponding to the first object they were tested with for each grasping pose. These user-independent models were subsequently fine-tuned using individual participant data containing one object per grasping pose.

Our system achieved an average microgesture recognition accuracy of 95.0% with non-deformable objects, which slightly decreased to 92.9% when interacting with deformable objects. The corresponding average false-positive rates were 0.2% and 0.3%, respectively. Note that the better performance observed with non-deformable

objects compared to the initial study can be attributed to two factors: (1) the base model benefits from being trained on a larger, combined dataset from both studies; (2) the evaluation utilized only one object per grasping pose, in contrast to the two objects used in the initial study. For the deformable objects, the lowest individual object accuracy was observed with the laminated foil pouches, primarily attributed to their inherent softness and significant shape variation along the grasping area. The pouches exhibited different deformation patterns with each grasp, and variations in grasping height further contributed to inconsistencies across data collection sessions. The second-lowest accuracy was recorded with the loofah, likely due to its extreme softness leading to high deformation variability during microgesture execution. Additionally, its slippery surface likely contributed to gradual slippage during data collection, particularly for participants with smaller hands. We acknowledge this inherent challenge posed by deformable objects; however, given the relatively small overall performance difference, we remain optimistic that the user experience in real-world applications will not be significantly impacted.

To facilitate further research in this domain, we released the whole dataset with 18 participants and 35 objects to the research community.

5.6 Post-study Wearability Survey Results

Given the identical apparatus used in both studies, we analyzed the wearing experience collectively. Firstly, none of the participants reported perceiving any sound from the device. This indicates that our frequency sweep is compatible with commercial speakers and microphones while remaining inaudible to users. Secondly, the participants generally found Grab-n-Go comfortable to wear (mean = 5.89, std = 1.18 on the Likert scale: 1 = extremely uncomfortable, 7 = extremely comfortable). Specifically, several participants described it as comfortable (P1-3, P1-4, P1-5, P1-12, P2-1, P2-2), fitting well (P1-7, P2-3, P2-7), and feeling like a normal wristband or watch (P1-1, P1-13, P2-7). However, one participant noted that the 3D-printed component felt slightly sharp (P2-8), and another experienced mild skin irritation from the rubber strap (P1-11). For future iterations and deployment, careful attention should be paid to the device's edges to eliminate any sharpness, and alternative strap materials should be explored to enhance user comfort for a wider range of skin sensitivities.

6 DISCUSSION

The performance of Grab-n-Go in the user study is promising as a proof of concept. The proposed methods can be further optimized for real-world applications, and we aim to discuss some of the key points for future improvements.

6.1 Hands-busy Interaction

With the increasing ubiquity of computing devices, traditional input modalities that rely heavily on explicit manual dexterity create conflicts between digital interaction and everyday physical tasks. This tension forces users to continually prioritize between their current hands-on activities and device manipulation. To address this challenge and support always-available input, researchers have investigated various microgestural interaction paradigms that can be seamlessly integrated into everyday scenarios where hands are already occupied [45–47]. Building upon these prior microgesture design strategies, we defined our own microgesture set for evaluation.

Recognizing the importance of hands-busy interaction, specifically input modalities that allow interaction even when hands are occupied, researchers have investigated various sensing modalities for microgesture recognition in such situations. Saponas et al. [39] utilized forearm EMG on an armband to recognize the pressing of four distinct fingers while holding a travel mug or bag. Rudolph et al. [38] proposed a capacitive sensing wristband capable of distinguishing dynamic object interactions like slide, rock, twiddle, squeeze, stretch, and tripod pinch,

with each interaction tied to a specific kind of object. VibAware [13] employed bio-acoustic sensing with a ring-wristband combination to detect thumb and index finger taps and swipes when grasping four specific 3D-printed props. SparseIMU [48] used distributed IMUs across finger joints to recognize six different microgestures while holding 12 distinct objects. However, the presence of objects introduces significant challenges, including diverse object geometries and potential occlusion, often leading to limitations in the variety of objects and microgestures investigated. While these prior works presented promising results within their specific contexts, the question of how well microgesture recognition with busy hands can generalize across a wide and diverse range of objects remains largely unanswered. This formed our primary design consideration (Sec. 3.1). By testing Grab-n-Go with 35 objects, encompassing both solid and deformable types, we verified the generalizability of Grab-n-Go, supporting its potential for future deployment in everyday life where users interact with a multitude of items. In addition, tackling our second design consideration regarding sensing technique and form factor (Sec. 3.2), Grab-n-Go achieves a promising performance while maintaining the lightweight watch-like form-factor. This preserves the potential integration with commercial smartwatches in the future. The comparison with prior work is summarized in Table 1.

6.2 Usage Scenarios

As discussed in the previous section, enabling hands-busy interaction across a variety of everyday objects is a crucial aspect of ubiquitous computing. With Grab-n-Go, we envisioned that the users could interact with their computers with subtle hand gestures even when their hands are occupied. This applies to short, fast, and discreet interaction scenarios. To illustrate a potential use case, consider Lisa, who is rushing to work and needs a coffee to stay alert. With her handbag in her left hand and a freshly purchased coffee in her right, she desires to switch her credit card for payment. Rather than struggling to find a place to set things down and swipe on her smartwatch, she can elegantly switch the credit card using the Pointer In microgesture on the coffee cup (Fig. 1).

In another scenario, Lisa is inspecting newly soldered prototypes and keeping relevant records on her laptop. To efficiently mark each soldering point after testing, she can use Wrist Right and Wrist Left microgestures to navigate between recording items. Thumb Tap can be used to mark a point as connected, while Pincky Out can indicate a disconnected point. When seeking a break, Lisa can grab her coffee and initiate music playback with a Pointer Tap microgesture, and use Wrist Left and Wrist Right to navigate between songs.

6.3 Hardware Design

To effectively capture signals surrounding the hand, we strategically placed two pairs of sensors on each side of the wrist. However, given that certain microgestures, such as those associated with the Spherical grasping pose, occur primarily on one side of the hand, and the opposite-side sensor can be occluded by in-hand objects, it is worth exploring the feasibility of using only a single speaker-microphone pair in future implementations. This approach could potentially reduce the system's complexity and improve its wearability, particularly for applications focused on a specific subset of microgestures.

6.4 Machine Learning Algorithm Comparison

Beyond the proposed ResNet-18 architecture, we sought to understand the efficacy of various machine learning algorithms when applied to our C-FMCW-based active acoustic data. We evaluated the performance of commonly used machine learning algorithms on the collected initial user study data (Sec. 5.2). For each model, we employed a consistent training and validation strategy: training on the first five data collection sessions and validating on the final session for each participant. The reported results represent the average performance across all participants.

Inspired by prior work with a similar goal [13, 38, 39, 48], we first evaluated with traditional machine learning methods from the Scikit-learn library [35]. We compared the performance of Linear Support Vector Classification

Table 2. Comparison of Different Machine Learning Algorithms

LinearSVC (Flatten)	**LinearSVC** (Haralick)	**LDA** (Flatten)	**LDA** (Haralick)	**Random Forest** (Flatten)	**Random Forest** (Haralick)	**TabPFN** (Haralick)	**CNN-LSTM**	**RepViT**	**FastViT**	**ResNet-18**
45.87%	72.33%	47.37%	73.54%	60.32%	71.67%	78.56%	85.60%	73.20%	85.11%	**86.63%**

(LinearSVC), Linear Discriminant Analysis (LDA), and Random Forest. Given our 2D image-like echo profiles, we initially used a flattened version of the echo profiles as input. As an alternative, we explored feature extraction. A pilot study investigated several common image feature extraction techniques, including color histograms, Haralick features, Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG). Based on the pilot study, Haralick features yielded the most promising results, and we subsequently used this method to generate an alternative input representation. Overall, training the LDA model with Haralick feature input produces the best results.

Subsequently, we explored deep learning methods. Given the relatively small size of our custom-collected dataset, we employed TabPFN [6], a pre-trained Transformer specifically designed for supervised classification on small tabular datasets. Leveraging its foundation model, which was trained on a large and diverse corpus of tabular data, our TabPFN model achieved an average accuracy of 78.56%.

With a focus on potential on-device deployment, we explored lightweight methods. Specifically, we adopted Reparameterized Vision Transformer (RepViT) [54], which combines CNN efficiency with Vision Transformer design principles, leveraging its ability to maintain high accuracy while significantly reducing computational demands. We also assessed FastViT [53], a hybrid architecture that strategically balances CNN and Transformer components to optimize both performance and processing speed. Our experiments demonstrated that RepViT achieved accuracy comparable to traditional methods, while FastViT surpassed them with an accuracy of 85.11%.

To better leverage the temporal dynamics inherent in the microgesture echo profiles, we developed a customized deep-learning network incorporating a CNN-LSTM encoder augmented with attention mechanisms and a fully connected classifier. This architecture yielded performance comparable to our ResNet-18 model, suggesting the temporal information captured by the LSTM provides a complementary representation of the microgestures.

Our evaluation revealed that our proposed Encoder-Decoder architecture leveraging a ResNet-18 backbone yielded the best results. However, it is important to highlight that the CNN-LSTM and FastViT achieved results comparable to those of this top-performing configuration. Given FastViT's significantly more lightweight nature in terms of computational complexity and memory footprint compared to the ResNet-18 model, it emerges as a highly promising candidate for future deployment scenarios, particularly on resource-constrained platforms such as mobile devices or even direct on-chip deployment.

6.5 Overfitting

In our evaluation, we observed a difference between the final training accuracy, which reached 100.0%, and the final testing accuracy of 92.0%. This 8.0% gap suggests a degree of overfitting, a phenomenon not unexpected given the inherent constraints of our dataset size and the complexity of the 30-class microgesture recognition task across 25 diverse objects. While this level of overfitting warrants consideration for future work, we believe the achieved 92.0% testing accuracy represents a strong level of real-world performance, potentially meeting the practical threshold for reliable and usable interaction within our intended application context. Nevertheless, we acknowledge that further research into mitigating this overfitting and enhancing the model's generalization capabilities could lead to even more robust and dependable performance in broader deployment scenarios.

6.6 Real-world Application

While this paper focuses on microgesture recognition with objects in hand, it is important to acknowledge that hands are not always occupied, and the system should not always be active. We plan to investigate methods for activating microgesture recognition only when necessary. One potential approach involves integrating a separate system to detect hand occupancy. Alternatively, introducing a unique activation gesture that is less likely to occur accidentally in daily life could also be considered.

As our evaluation was conducted in a controlled in-lab setting, the robustness of the microgesture sets in real-world environments remains to be explored. Some microgestures may be commonly used for other purposes, potentially leading to the accidental activation of unintended functions in daily life. Future research should investigate the system's robustness, following the approach outlined in SoloFinger [45], to ensure its suitability for real-world deployment.

Considering the two-step fine-tuning scheme's implications for usability in future deployments, we envision leveraging a large, shared base dataset for the initial training of a foundational model. This approach would streamline the user experience, requiring new users to collect only a single set of personalized data during their initial device setup, akin to the familiar process of registering face ID on smartphones or configuring new gestural inputs on augmented/virtual reality devices. This minimal initial effort would offer a streamlined and user-friendly onboarding process.

6.7 The Impact of Object Selections

While we have demonstrated that Grab-n-Go successfully recognizes microgestures across 35 different objects as a proof of concept for a research prototype, the range of objects people interact with every day goes far beyond this number. We plan to further investigate the system's ability to recognize microgestures on unseen objects that share similar grasping poses. This will include assessing whether the current model can accurately classify microgestures on objects not included in our existing dataset. While we presented the object-independent results (Sec. 5.4.4), further exploration with more diverse data within the same grasping pose category is needed. Ultimately, our goal is to enable Grab-n-Go to support microgesture recognition across any arbitrary object.

Since we leverage ultrasonic range for our active acoustic sensing method, external noises generally do not interfere with the received acoustic signals. However, when certain materials are taped or scratched, they can produce high-frequency signals. By incorporating a diverse range of materials of objects into our dataset, we effectively mitigated the impact of these signals.

6.8 Microgesture Set

Despite providing participants with two seconds to perform each microgesture, the execution speed varied significantly. Additionally, the level of exaggeration in microgestures also differed. However, training the base model on data from all participants effectively addressed these challenges, demonstrating Grab-n-Go's ability to handle diverse microgesture styles.

Given the variation in participant hand sizes and object dimensions, the manner in which objects were held also differed significantly. For instance, while a typical grasping pose for Spherical objects involves an arched palm, a participant with smaller hands found it challenging to grasp the jar lid, leading to a flatter palm. Despite these variations, Grab-n-Go effectively adapted to these diverse holding styles.

Although Grab-n-Go successfully recognized 30 microgestures, the specific number required may vary depending on the application. For instance, 3 microgestures per grasping pose might be sufficient for quick tasks on commercial smartwatches, like answering calls or muting the device. Moreover, focusing on microgestures associated with a single grasping pose, such as pen interaction with the Tip microgesture when using an Apple

Pencil, could be a potential use case. By reducing the microgesture set, we anticipate further improvements in performance.

Our evaluation showcased Grab-n-Go's capability to support a comprehensive set of 30 distinct microgestures. While the design of this set was informed by prior research and considerations for ease of execution across various object shapes and weights, we recognize that such a large set could potentially impose a significant cognitive load on users for memorization in practical applications. The primary goal of our evaluation was to rigorously investigate the fundamental recognition capabilities of Grab-n-Go across a wide range of objects, hence our decision to test this extensive microgesture set as a proof of concept. For real-world deployment scenarios, we plan to collaborate closely with user experience researchers to carefully curate a more streamlined and intuitive microgesture set that balances functionality with ease of memorization and use.

6.9 Limitation

While our evaluation was conducted in a controlled lab setting, some participants exhibited movement due to the extended study duration. However, the overall movement is limited. We plan to explore Grab-n-Go's performance during more dynamic activities, such as walking, to gain a deeper understanding of its capabilities in real-world scenarios.

Additionally, we observed that certain objects, when held in the hand, can significantly obstruct acoustic signals, hindering the recognition of finger microgestures. For instance, when holding a pillow using the Palmar grasping pose, the softness of the pillow can cause the hand, including the sensors, to sink into the material, preventing the propagation of in-air acoustic signals. As a result, the corresponding echo profiles lacked distinct patterns, making microgesture detection unreliable. Grab-n-Go may not work well on these objects, which is another limitation of our proposed system.

In terms of the occlusion, Grab-n-Go's capability will also be constrained by the covering of the clothes. It is natural that sometimes the watch will be covered by the sleeve. However, if being covered by the sleeve, Grab-n-Go will suffer from the signals being blocked by the sleeve and can not capture the information from the desired area.

7 CONCLUSION

In this paper, we introduce Grab-n-Go, a wristband that enables robust recognition of hand microgestures when hands are occupied with objects. Leveraging active acoustic sensing, Grab-n-Go effectively identifies 30 microgestures with an average accuracy of 92% across a diverse set of 25 objects. The system was evaluated through a user study with 10 participants, involving the collection of 14,400 microgesture instances. A follow-up study with an additional 8 participants further expanded our dataset by collecting 5,760 more microgesture instances, specifically to validate the system's robustness against more challenging, deformable objects. Overall, Grab-n-Go showcases its effectiveness in enabling microgesture recognition across a wide range of everyday objects, providing a seamless solution for always-available input.

Acknowledgments

This project was supported by National Science Foundation Grant No. 2239569. We want to thank our colleagues at SciFi Lab for their invaluable support, all participants for their generous contributions to the user study, and the reviewers for their insightful feedback.

References

[1] Min Chen, Jingyu Ouyang, Aijia Jian, Jia Liu, Pan Li, Yixue Hao, Yuchen Gong, Jiayu Hu, Jing Zhou, Rui Wang, et al. 2022. Imperceptible, designable, and scalable braided electronic cord. *Nature Communications* 13, 1 (2022), 7097.

[2] Yu Du, Yongkang Wong, Wenguang Jin, Wentao Wei, Yu Hu, Mohan Kankanhalli, and Weidong Geng. 2017. Semi-Supervised Learning for Surface EMG-based Gesture Recognition. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization. doi:10.24963/ijcai.2017/225

[3] 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

[4] Jun Gong, Olivia Seow, Cedric Honnet, Jack Forman, and Stefanie Mueller. 2021. MetaSense: Integrating Sensing Capabilities into Mechanical Metamaterial. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 1063–1073. doi:10.1145/3472749.3474806

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[6] Noah Hollmann, Samuel Müller, Katharina Eggensperger, and Frank Hutter. 2022. Tabpfn: A transformer that solves small tabular classification problems in a second. *arXiv preprint arXiv:2207.01848* (2022).

[7] Chien-Ti Hsiao, Pei-Shin Hwang, Polly Huang, Kate Ching-Ju Lin, and Ling-Jyh Chen. 2024. Demo Abstract: Polyband – A Carbon Polymer Wristband for Hand Gesture Recognition. In *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems* (Istanbul, Turkiye) *(SenSys '23)*. Association for Computing Machinery, New York, NY, USA, 480–481. doi:10.1145/3625687.3628390

[8] 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

[9] Yoshihiro Kawahara, Steve Hodges, Benjamin S. Cook, Cheng Zhang, and Gregory D. Abowd. 2013. Instant inkjet circuits: lab-based inkjet printing to support rapid prototyping of UbiComp devices. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 363–372. doi:10.1145/2493432.2493486

[10] Frederic Kerber, Michael Puhl, and Antonio Krüger. 2017. User-Independent Real-Time Hand Gesture Recognition Based on Surface Electromyography. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 36, 7 pages. doi:10.1145/3098279.3098553

[11] Daehwa Kim and Chris Harrison. 2022. EtherPose: Continuous Hand Pose Tracking with Wrist-Worn Antenna Impedance Characteristic Sensing. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 58, 12 pages. doi:10.1145/3526113.3545665

[12] David Kim, Otmar Hilliges, Shahram Izadi, Alex D Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. 167–176.

[13] Jina Kim, Minyung Kim, Woo Suk Lee, and Sang Ho Yoon. 2023. VibAware: Context-Aware Tap and Swipe Gestures Using Bio-Acoustic Sensing. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction* (Sydney, NSW, Australia) *(SUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 6, 12 pages. doi:10.1145/3607822.3614544

[14] Jiwan Kim, Jiwan Son, and Ian Oakley. 2025. Cross, Dwell, or Pinch: Designing and Evaluating Around-Device Selection Methods for Unmodified Smartwatches. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 604, 11 pages. doi:10.1145/3706598.3714308

[15] Pin-Sung Ku, Qijia Shao, Te-Yen Wu, Jun Gong, Ziyan Zhu, Xia Zhou, and Xing-Dong Yang. 2020. Threadsense: Locating touch on an extremely thin interactive thread. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[16] Alexander Kyu, Hongyu Mao, Junyi Zhu, Mayank Goel, and Karan Ahuja. 2024. EITPose: Wearable and Practical Electrical Impedance Tomography for Continuous Hand Pose Estimation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–10.

[17] Chi-Jung Lee, Ruidong Zhang, Devansh Agarwal, Tianhong Catherine Yu, Vipin Gunda, Oliver Lopez, James Kim, Sicheng Yin, Boao Dong, Ke Li, et al. 2024. Echowrist: Continuous hand pose tracking and hand-object interaction recognition using low-power active acoustic sensing on a wristband. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–21.

[18] Hong Li, Shishir Chawla, Richard Li, Sumeet Jain, Gregory D Abowd, Thad Starner, Cheng Zhang, and Thomas Plötz. 2018. Wristwash: towards automatic handwashing assessment using a wrist-worn device. In *Proceedings of the 2018 ACM international symposium on wearable computers*. 132–139.

[19] Ke Li, Devansh Agarwal, Ruidong Zhang, Vipin Gunda, Tianjun Mo, Saif Mahmud, Boao Chen, François Guimbretière, and Cheng Zhang. 2024. SonicID: User Identification on Smart Glasses with Acoustic Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 4, Article 169 (Nov. 2024), 27 pages. doi:10.1145/3699734

[20] Ke Li, Ruidong Zhang, Siyuan Chen, Boao Chen, Mose Sakashita, Francois Guimbretiere, and Cheng Zhang. 2024. EyeEcho: Continuous and Low-power Facial Expression Tracking on Glasses. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 319, 24 pages. doi:10.1145/3613904.3642613

[21] Ke Li, Ruidong Zhang, Bo Liang, François Guimbretière, and Cheng Zhang. 2022. Eario: A low-power acoustic sensing earable for continuously tracking detailed facial movements. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–24.

[22] Hyunchul Lim, Nam Anh Dang, Dylan Lee, Tianhong Catherine Yu, Jane Lu, Franklin Mingzhe Li, Yiqi Jin, Yan Ma, Xiaojun Bi, François Guimbretière, and Cheng Zhang. 2025. SpellRing: Recognizing Continuous Fingerspelling in American Sign Language using a Ring. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 67, 17 pages. doi:10.1145/3706598.3713721

[23] Yang Liu, Chengdong Lin, and Zhenjiang Li. 2021. WR-Hand: Wearable Armband Can Track User's Hand. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 118 (sep 2021), 27 pages. doi:10.1145/3478112

[24] Yilin Liu, Shijia Zhang, and Mahanth Gowda. 2021. NeuroPose: 3D Hand Pose Tracking Using EMG Wearables. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 1471–1482. doi:10.1145/3442381.3449890

[25] C.L. MacKenzie and T. Iberall. 1994. *The Grasping Hand.* Elsevier Science. https://books.google.com/books?id=V9G5Yd46VlEC

[26] Saif Mahmud, Devansh Agarwal, Ashwin Ajit, Qikang Liang, Thalia Viranda, Francois Guimbretiere, and Cheng Zhang. 2024. MunchSonic: Tracking Fine-grained Dietary Actions through Active Acoustic Sensing on Eyeglasses. In *Proceedings of the 2024 ACM International Symposium on Wearable Computers* (Melbourne VIC, Australia) *(ISWC '24)*. Association for Computing Machinery, New York, NY, USA, 96–103. doi:10.1145/3675095.3676619

[27] Saif Mahmud, Ke Li, Guilin Hu, Hao Chen, Richard Jin, Ruidong Zhang, François Guimbretière, and Cheng Zhang. 2023. Posesonic: 3d upper body pose estimation through egocentric acoustic sensing on smartglasses. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (2023), 1–28.

[28] Saif Mahmud, Vineet Parikh, Qikang Liang, Ke Li, Ruidong Zhang, Ashwin Ajit, Vipin Gunda, Devansh Agarwal, Francois Guimbretiere, and Cheng Zhang. 2024. ActSonic: Recognizing Everyday Activities from Inaudible Acoustic Wave Around the Body. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 4, Article 183 (Nov. 2024), 32 pages. doi:10.1145/3699752

[29] 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

[30] Alex Olwal, Jon Moeller, Greg Priest-Dorman, Thad Starner, and Ben Carroll. 2018. I/O Braid: Scalable touch-sensitive lighted cords using spiraling, repeating sensing textiles and fiber optics. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 485–497.

[31] Alex Olwal, Thad Starner, and Gowa Mainini. 2020. E-Textile microinteractions: Augmenting twist with flick, slide and grasp gestures for soft electronics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[32] Gianpaolo Palma, Narges Pourjafarian, Jürgen Steimle, and Paolo Cignoni. 2024. Capacitive Touch Sensing on General 3D Surfaces. *ACM Trans. Graph.* 43, 4, Article 103 (July 2024), 20 pages. doi:10.1145/3658185

[33] Vineet Parikh, Saif Mahmud, Devansh Agarwal, Ke Li, François Guimbretière, and Cheng Zhang. 2024. EchoGuide: Active Acoustic Guidance for LLM-Based Eating Event Analysis from Egocentric Videos. In *Proceedings of the 2024 ACM International Symposium on Wearable Computers* (Melbourne VIC, Australia) *(ISWC '24)*. Association for Computing Machinery, New York, NY, USA, 40–47. doi:10.1145/3675095.3676611

[34] Taiwoo Park, Jinwon Lee, Inseok Hwang, Chungkuk Yoo, Lama Nachman, and Junehwa Song. 2011. E-Gesture: a collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems* (Seattle, Washington) *(SenSys '11)*. Association for Computing Machinery, New York, NY, USA, 260–273. doi:10.1145/2070942.2070969

[35] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* 12 (2011), 2825–2830.

[36] Narjes Pourjafarian, Marion Koelle, Fjolla Mjaku, Paul Strohmeier, and Jürgen Steimle. 2022. Print-A-Sketch: A Handheld Printer for Physical Sketching of Circuits and Sensors on Everyday Surfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 270, 17 pages. doi:10.1145/3491102.3502074

[37] Sumit Raurale, John McAllister, and Jesus Martinez del Rincon. 2018. Emg Acquisition and Hand Pose Classification for Bionic Hands from Randomly-Placed Sensors. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Calgary, AB, Canada). IEEE Press, 1105–1109. doi:10.1109/ICASSP.2018.8462409

[38] Julius Cosmo Romeo Rudolph, David Holman, Bruno De Araujo, Ricardo Jota, Daniel Wigdor, and Valkyrie Savage. 2022. Sensing hand interactions with everyday objects by profiling wrist topography. In *Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–14.

[39] T. Scott Saponas, Desney S. Tan, Dan Morris, Ravin Balakrishnan, Jim Turner, and James A. Landay. 2009. Enabling always-available input with muscle-computer interfaces. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology* (Victoria, BC, Canada) *(UIST '09)*. Association for Computing Machinery, New York, NY, USA, 167–176. doi:10.1145/1622176.1622208

[40] Munehiko Sato, Ivan Poupyrev, and Chris Harrison. 2012. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 483–492. doi:10.1145/2207676.2207743

[41] Valkyrie Savage, Xiaohan Zhang, and Björn Hartmann. 2012. Midas: fabricating custom capacitive touch sensors to prototype interactive objects. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (Cambridge, Massachusetts, USA) *(UIST '12)*. Association for Computing Machinery, New York, NY, USA, 579–588. doi:10.1145/2380116.2380189

[42] G. Schlesinger. 1919. *Der mechanische Aufbau der künstlichen Glieder*. Springer Berlin Heidelberg, Berlin, Heidelberg, 321–661. doi:10.1007/978-3-662-33009-8_13

[43] Martin Schmitz, Mohammadreza Khalilbeigi, Matthias Balwierz, Roman Lissermann, Max Mühlhäuser, and Jürgen Steimle. 2015. Capricate: A fabrication pipeline to design and 3D print capacitive touch sensors for interactive objects. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 253–258.

[44] Fereshteh Shahmiri, Chaoyu Chen, Anandghan Waghmare, Dingtian Zhang, Shivan Mittal, Steven L Zhang, Yi-Cheng Wang, Zhong Lin Wang, Thad E Starner, and Gregory D Abowd. 2019. Serpentine: A self-powered reversibly deformable cord sensor for human input. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[45] Adwait Sharma, Michael A Hedderich, Divyanshu Bhardwaj, Bruno Fruchard, Jess McIntosh, Aditya Shekhar Nittala, Dietrich Klakow, Daniel Ashbrook, and Jürgen Steimle. 2021. SoloFinger: Robust microgestures while grasping everyday objects. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–15.

[46] Adwait Sharma, Alexander Ivanov, Frances Lai, Tovi Grossman, and Stephanie Santosa. 2024. GraspUI: Seamlessly Integrating Object-Centric Gestures within the Seven Phases of Grasping. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 1275–1289. doi:10.1145/3643834.3661551

[47] Adwait Sharma, Joan Sol Roo, and Jürgen Steimle. 2019. Grasping Microgestures: Eliciting Single-hand Microgestures for Handheld Objects. 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:10.1145/3290605.3300632

[48] Adwait Sharma, Christina Salchow-Hömmen, Vimal Suresh Mollyn, Aditya Shekhar Nittala, Michael A. Hedderich, Marion Koelle, Thomas Seel, and Jürgen Steimle. 2023. SparseIMU: Computational Design of Sparse IMU Layouts for Sensing Fine-grained Finger Microgestures. *ACM Trans. Comput.-Hum. Interact.* 30, 3, Article 39 (jun 2023), 40 pages. doi:10.1145/3569894

[49] Ashwin De Silva, Malsha V. Perera, Kithmin Wickramasinghe, Asma M. Naim, Thilina Dulantha Lalitharatne, and Simon L. Kappel. 2020. Real-Time Hand Gesture Recognition Using Temporal Muscle Activation Maps of Multi-Channel Semg Signals. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1299–1303. doi:10.1109/ICASSP40776.2020.9054227

[50] Rujia Sun, Xiaohe Zhou, Benjamin Steeper, Ruidong Zhang, Sicheng Yin, Ke Li, Shengzhang Wu, Sam Tilsen, Francois Guimbretiere, and Cheng Zhang. 2023. EchoNose: Sensing Mouth, Breathing and Tongue Gestures inside Oral Cavity using a Non-contact Nose Interface. In *Proceedings of the 2023 ACM International Symposium on Wearable Computers* (Cancun, Quintana Roo, Mexico) *(ISWC '23)*. Association for Computing Machinery, New York, NY, USA, 22–26. doi:10.1145/3594738.3611358

[51] Carlos E. Tejada, Raf Ramakers, Sebastian Boring, and Daniel Ashbrook. 2020. AirTouch: 3D-printed Touch-Sensitive Objects Using Pneumatic Sensing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3313831.3376136

[52] 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

[53] Pavan Kumar Anasosalu Vasu, James Gabriel, Jeff Zhu, Oncel Tuzel, and Anurag Ranjan. 2023. Fastvit: A fast hybrid vision transformer using structural reparameterization. In *Proceedings of the IEEE/CVF international conference on computer vision*. 5785–5795.

[54] Ao Wang, Hui Chen, Zijia Lin, Jungong Han, and Guiguang Ding. 2024. Repvit: Revisiting mobile cnn from vit perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15909–15920.

[55] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW Based Contactless Respiration Detection Using Acoustic Signal. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 170 (Jan. 2018), 20 pages. doi:10.1145/3161188

[56] Michael Wessely, Ticha Sethapakdi, Carlos Castillo, Jackson C. Snowden, Ollie Hanton, Isabel P. S. Qamar, Mike Fraser, Anne Roudaut, and Stefanie Mueller. 2020. Sprayable User Interfaces: Prototyping Large-Scale Interactive Surfaces with Sensors and Displays. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3313831.3376249

[57] 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*. 1147–1160.

[58] Te-Yen Wu, Shutong Qi, Junchi Chen, MuJie Shang, Jun Gong, Teddy Seyed, and Xing-Dong Yang. 2020. Fabriccio: Touchless Gestural Input on Interactive Fabrics. Association for Computing Machinery, New York, NY, USA, 1–14.

[59] Te-Yen Wu, Zheer Xu, Xing-Dong Yang, Steve Hodges, and Teddy Seyed. 2021. Project Tasca: Enabling Touch and Contextual Interactions with a Pocket-Based Textile Sensor. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[60] Hui-Shyong Yeo, Erwin Wu, Juyoung Lee, Aaron Quigley, and Hideki Koike. 2019. Opisthenar: Hand Poses and Finger Tapping Recognition by Observing Back of Hand Using Embedded Wrist Camera. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 963–971. doi:10.1145/3332165.3347867

[61] Tianhong Catherine Yu, Riku Arakawa, James McCann, and Mayank Goel. 2023. uKnit: A Position-Aware Reconfigurable Machine-Knitted Wearable for Gestural Interaction and Passive Sensing using Electrical Impedance Tomography. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[62] Tianhong Catherine Yu, Guilin Hu, Ruidong Zhang, Hyunchul Lim, Saif Mahmud, Chi-Jung Lee, Ke Li, Devansh Agarwal, Shuyang Nie, Jinseok Oh, François Guimbretière, and Cheng Zhang. 2024. Ring-a-Pose: A Ring for Continuous Hand Pose Tracking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 4, Article 189 (Nov. 2024), 30 pages. doi:10.1145/3699741

[63] Cheng Zhang, AbdelKareem Bedri, Gabriel Reyes, Bailey Bercik, Omer T Inan, Thad E Starner, and Gregory D Abowd. 2016. TapSkin: Recognizing on-skin input for smartwatches. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*. 13–22.

[64] Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Sumeet Jain, Yiming Pu, Sinan Hersek, Kent Lyons, Kenneth A Cunefare, Omer T Inan, and Gregory D Abowd. 2017. Soundtrak: Continuous 3d tracking of a finger using active acoustics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 1–25.

[65] Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Ruichen Meng, Sumeet Jain, Yizeng Han, Xinyu Li, Kenneth Cunefare, Thomas Ploetz, Thad Starner, Omer Inan, and Gregory D. Abowd. 2018. FingerPing: Recognizing Fine-grained Hand Poses using Active Acoustic On-body Sensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Montreal QC</city>, <country>Canada</country>, </conf-loc>) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3173574.3174011

[66] Cheng Zhang, Junrui Yang, Caleb Southern, Thad E Starner, and Gregory D Abowd. 2016. WatchOut: extending interactions on a smartwatch with inertial sensing. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers*. 136–143.

[67] Ruidong Zhang, Hao Chen, Devansh Agarwal, Richard Jin, Ke Li, François Guimbretière, and Cheng Zhang. 2023. HPSpeech: Silent Speech Interface for Commodity Headphones. In *Proceedings of the 2023 ACM International Symposium on Wearable Computers* (Cancun, Quintana Roo, Mexico) *(ISWC '23)*. Association for Computing Machinery, New York, NY, USA, 60–65. doi:10.1145/3594738.3611365

[68] Ruidong Zhang, Ke Li, Yihong Hao, Yufan Wang, Zhengnan Lai, François Guimbretière, and Cheng Zhang. 2023. EchoSpeech: Continuous Silent Speech Recognition on Minimally-obtrusive Eyewear Powered by Acoustic Sensing. 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 852, 18 pages. doi:10.1145/3544548.3580801

[69] Yang Zhang and Chris Harrison. 2015. Tomo: Wearable, Low-Cost Electrical Impedance Tomography for Hand Gesture Recognition. 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, 167–173. doi:10.1145/2807442.2807480

[70] Yang Zhang, Gierad Laput, and Chris Harrison. 2017. Electrick: Low-cost touch sensing using electric field tomography. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 1–14.

[71] Yang Zhang, Robert Xiao, and Chris Harrison. 2016. Advancing Hand Gesture Recognition with High Resolution Electrical Impedance Tomography. 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, 843–850. doi:10.1145/2984511.2984574

[72] Joshua Z Zheng, Sara De La Rosa, and Aaron M Dollar. 2011. An investigation of grasp type and frequency in daily household and machine shop tasks. In *2011 IEEE international conference on robotics and automation*. IEEE, 4169–4175.