SoloFinger: Robust Microgestures while Grasping Everyday Objects

Adwait Sharma
Saarland University, Saarland Informatics Campus
Saarbrücken, Germany
sharma@cs.uni-saarland.de

Michael A. Hedderich
Saarland University, Saarland Informatics Campus
Saarbrücken, Germany
mhedderich@lsv.uni-saarland.de

Divyanshu Bhardwaj
Saarland University, Saarland Informatics Campus
Saarbrücken, Germany
s8dibhar@stud.uni-saarland.de

Bruno Fruchard
Saarland University, Saarland Informatics Campus
Saarbrücken, Germany
fruchard@cs.uni-saarland.de

Jess McIntosh
University of Copenhagen
Copenhagen, Denmark
jemc@di.ku.dk

Aditya Shekhar Nittala
Saarland University, Saarland Informatics Campus
Saarbrücken, Germany
nittala@cs.uni-saarland.de

Dietrich Klakow
Saarland University, Saarland Informatics Campus
Saarbrücken, Germany
dietrich.klakow@lsv.uni-saarland.de

Daniel Ashbrook
University of Copenhagen
Copenhagen, Denmark
dan@di.ku.dk

Jürgen Steimle
Saarland University, Saarland Informatics Campus
Saarbrücken, Germany
steimle@cs.uni-saarland.de

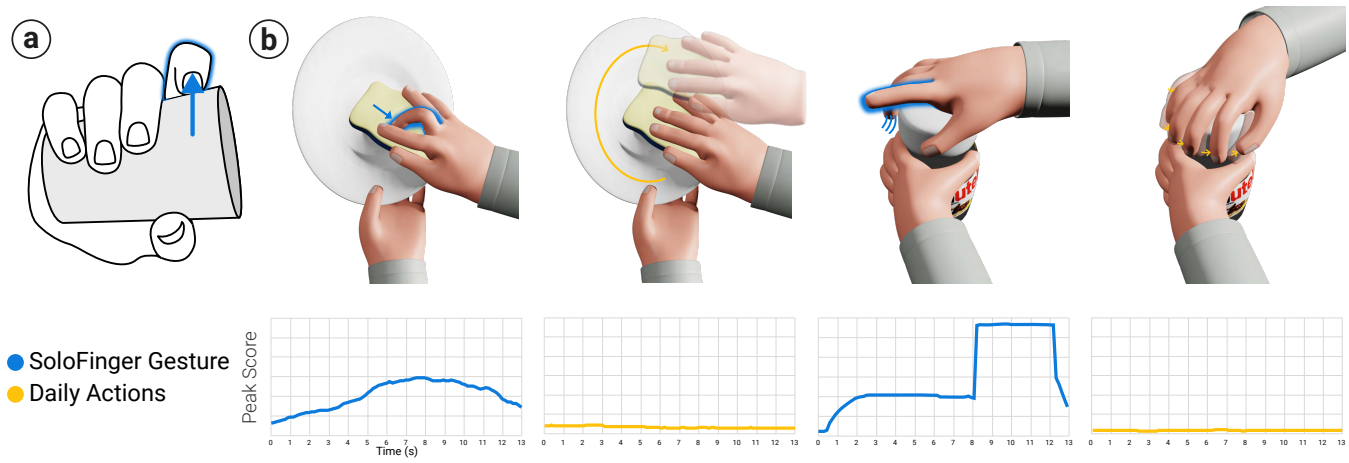


Figure 1: (a) The SoloFinger concept: while grasping an object, one can perform a single-finger microgesture while other fingers stay idle. (b) These easy and rapid-to-perform gestures exhibit a distinct movement signature, which is not present during daily actions. This yields a robust gestural input compatible with versatile object geometries and actions.

ABSTRACT

Using microgestures, prior work has successfully enabled gestural interactions while holding objects. Yet, these existing methods are prone to false activations caused by natural finger movements while holding or manipulating the object. We address this issue with *SoloFinger*, a novel concept that allows design of microgestures that are robust against movements that naturally occur during primary activities. Using a data-driven approach, we establish that single-finger movements are rare in everyday hand-object actions and infer a single-finger input technique resilient to false activation. We demonstrate this concept's robustness using a white-box classifier on a pre-existing dataset comprising 36 everyday hand-object actions. Our findings validate that simple SoloFinger gestures can relieve the need for complex finger configurations or delimiting gestures and that SoloFinger is applicable to diverse hand-object actions. Finally, we demonstrate SoloFinger's high performance on commodity hardware using random forest classifiers.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI '21, May 8–13, 2021, Yokohama, Japan

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8096-6/21/05.
https://doi.org/10.1145/3411764.3445197

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Interaction techniques**; **Gestural input**.

KEYWORDS

microgesture; grasping; everyday objects; false activation

ACM Reference Format:
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 *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3411764.3445197

1 INTRODUCTION

Recent work has identified microgestures as a promising means for enabling interaction while holding everyday objects, such as a pen, a hammer, or a steering wheel [38, 58, 72, 74]. These small and rapid gestures (e.g., a tapping or a sliding movement) are one-handed and can be performed directly on objects, notably with the same hand that is holding the object. For this property they have been referred to as *grasping microgestures* [58]. Such gestures might be used to replace the wake word in virtual assistant spaces (e.g., "Alexa" or "Hey Google") [48], but can also offer benefits in a myriad of applications, ranging from controlling mobile devices when on the move and hands are busy, to controlling systems in healthcare contexts [8].

While performing quick and convenient gestures on handheld objects is compelling, gestures risk conflicting with finger movements that might occur when adjusting ones grip or manipulating the object. As a result, a gesture recognizer might misinterpret natural finger movements as an intentional input gesture and trigger an unintended command; namely, a *false activation*. This is a challenging problem for designing and deploying microgestural interfaces.

To the best of our knowledge, however, there is no prior work that systematically investigates robust gestures to perform while holding everyday objects. Previous work presented robust gestures for specific devices, such as smartphones [35, 50, 52], tablets [57], and smart pens [23]. While these approaches perform well in their specific contexts, a critical problem is their device-specific behavior. For example, a flipping gesture might be suitable for smartphones [52], but would not work while holding a coffee cup. Other work proposed using a robust delimiter gesture that must be performed before doing the actual gesture [53]. Yet another approach could involve complicated movements that are unlikely to happen in everyday actions, such as a specific movement sequence, specific timing, or specific finger combination. In contrast, the goal of our work is to move beyond device-specific gestures and ensure compatibility with a wide range of grasps and everyday actions, while avoiding the cognitive overhead of separate delimiters or complicated-to-perform gestures.

We present *SoloFinger*, a concept addressing the problem of false activation while grasping everyday objects. The main finding of this work is that simple finger movements that are rapid, easy and elegant to perform can indeed function as robust microgestures while holding objects. We demonstrate that this holds true for diverse grasps, object geometries and everyday actions.

SoloFinger leverages the insight that when holding and manipulating objects fingers tend either to be static or to have multiple fingers move concurrently. Consequently a movement of a single finger while other fingers remain idle stands out from everyday hand motions during object interaction (see Figure 1a and 1b). This opens up a space for performing robust yet easy-to-perform gestures on everyday objects that do not provide interactive input options of their own. To methodologically validate this in the context of hand actions and gesture design, we conducted extensive user studies and performed a series of data-driven analyses. As a result, we recommend 7 types of SoloFinger gestures performed with the thumb, index, or middle finger, offering overall 21 interaction options.

We start by introducing the overall concept and demonstrate its principled feasibility. Specifically, we introduce a metric that highlights the distinct behaviour of SoloFinger microgestures from everyday finger movements. We also show the feasibility of performing these gestures on diverse objects and with diverse grasps. Next, we empirically investigate in more detail the robustness of SoloFinger microgestures to false activation. We collected 7,488 gesture trials. We systematically analyze these gestures as well as a pre-existing dataset comprising 933 trials with daily hand-object actions. To produce findings that generalize beyond a specific classifier implementation and can be interpreted by humans, we opted for a simple white-box classifier, based on thresholding. The results show that SoloFinger microgestures performed on 36 objects can be recognized with an average precision of 100% and recall of 88% (SD = 7) over three primitive gestures. We also show the technique's resilience to false activation on the held-out dataset, which triggers false activation in only 51 out of 933 trials of actions performed with 36 objects. Notably, no false activation was found for 23 actions, while most cases of false activation occurred on extremely deformable or very small objects.

Finally, we demonstrate a proof-of-concept system with a commercially available virtual reality glove and a random forest classifier. This implementation can detect 7 types of SoloFinger microgestures performed with the thumb, index or middle finger. Classification results without knowledge of the held object show an overall accuracy of 86% with a very low number of false activations (17 out of 800 trials). When the held object is known, the accuracy further increases to 89%, without any false activation in the collected dataset.

In summary, our work makes the following contributions:

- **Concept**: We propose the concept of *SoloFinger microgestures*—where one finger moves intentionally while the others remain idle—as a means for rapid and convenient gestural input while holding an object that is robust to false activations.
- **Concept validation**: To validate our concept, we perform a systematic data-driven analysis. We show that SoloFinger gestures differ considerably from everyday hand motions and, using a white-box classifier, demonstrate the concept's principled feasibility for robust gesture detection.

- **Proof-of-concept system**: We demonstrate the real-world feasibility of our presented concept by implementing an end-to-end system with off-the-shelf hardware.
- **Dataset**: We release two datasets of SoloFinger gestures performed by a total of 21 participants and captured using an OptiTrack optical motion capture system and a virtual reality data glove for 36 and 5 actions, respectively. This fills a gap in the existing literature by providing data about finger gestures while grasping. The datasets are available at: *https://hci.cs.uni-saarland.de/research/solofinger/*

2 RELATED WORK

Our work relates to prior work in enabling interaction with handheld objects, approaches to avoid false activations, and gesture recognition.

2.1 Microgestures on Objects

The vast majority of work on touch gestures is based on the assumption that at least one of the user's hands is free for interacting with the touch surface. A relatively small amount of work has investigated settings in which the user's hands and fingers are busy with another task. Heo et al. [22] presented a taxonomy of such constrained scenarios, which illustrated that each type of constraint affords a different set of gestures.

As discussed by Norman [43], different physical objects provide different affordances; therefore, it is important to consider grasp while designing microinteractions. Prior literature has identified over 70 grasp types covering different perspectives, on the basis of tool design [55], task requirement [40], specific domain [9], and several other factors [14]. A detailed survey of different grasp taxonomies can be found in [36].

Previous work investigated how intentional interactions should be designed when the hands are occupied. Fitzmaurice et al. contributed seminal work with the concept of Graspable User Interfaces [15]. Graspables [65] presented grasp-recognition as an input medium. Wimmer presented the GRASP model to show how a grasp conveys meaningful contextual information [71]. Wolf et al. assessed the ergonomic factors and suitability of microinteractions in an expert study with 3 objects (steering wheel, pen and cash card) [72]. Other work has identified gestures focusing on specific objects and geometries, such as steering wheels [3], bike handles [11], textile-based cords [44], and pens [38]. More recently, Sharma et al. [58] focused on grasping microgestures through an elicitation study and contributed microgestures for six grasp types and 12 objects of various geometry. None of this work, however, has systematically investigated how to design microgestures that are robust against false activations. To address this gap, we present an empirically validated method to create comfortable yet robust microgestures for handheld objects that avoid false activations.

Various sensing technologies have demonstrated that capturing microgestures with everyday objects is technically feasible, e.g., [2, 21, 32, 37, 51, 56, 60, 63, 76]. We discuss the potential of several sensing techniques to realize our concept in Section 8.1.

2.2 False Activation During Gestural Input

Gesture detection errors can be classified in two categories: false positive errors, which relate to triggering unintended actions, and false negative errors, where a recognition system fails to identify the intended gesture. In HCI, both these errors result in user frustration and have direct implications on the adoption of a particular technology [69].

Some previous work has focused on designing explicit *delimiter gestures* to avoid false activations. These gestures are significantly different from non-intentional actions and are therefore robust to unintended input. One common application for such delimiter gestures is switching between a gesture detection mode and another mode in which the gesture recognizer is not active. Through our literature review, we found two main categories of delimiter action for handheld objects: *bimanual gestures*, where the trigger action is performed with the non-dominant hand and, subsequently, the command is performed with the dominant hand [53]; and *device-specific* trigger actions like DoubleFlip [52] that require a large rotation, or Active Edge [50] that uses squeezing to detect intentional action on phone devices. WristRotate [27] presented a wrist rotation technique as a delimiter for smartwatches. Recently, BlyncSync [67] used multi-modal touch and blink gestures on smartwatches. A drawback of any delimiter action is the disruption in the user's workflow: the user must first perform the delimiter, and then the intended gesture. Also, compound finger movements can be formed using Rhythmic microgestures [16]. More closely related to our approach, Le et al. [31] explored reachability and unintended input for a specific grasp type with different phone sizes. From a technical standpoint, some machine-learning approaches have been utilized previously to reduce false positive rates [10, 29]. However, these techniques are applied after the design process of gestures. Secondly, they do not consider false positive reduction while performing the gestures during hand-object interactions. Other efforts include engaging end-users in the design process of gestures to reduce the risk of confusing gestures with natural movement [58, 74]. Our contribution is to employ a data-driven approach to validate a set of intentional gestures, which are resilient to false activation on un-instrumented everyday objects and various grasp types.

Our approach for designing gestures is inspired by work from Kawahata et al. [26], Magic 2.0 [28], and Gesture On [35], which compare gestures against a database of everyday actions in order to identify the ones most robust to false activations. We take a more general approach by investigating a large number of grasp types to design a gesture language compatible with many different objects. Our approach results in unique single-finger motions with the same hand involved in holding the object.

2.3 Gesture Recognition Techniques

Recognizing gestures from sensor data can be framed as a classification task for a machine-learning model. In this work, we opted for two different approaches: *white-box* and *black-box*. A white-box model is a machine-learning technique that can be easily interpreted by a human. The advantage of such an interpretable and transparent system is that one can understand the decision process of the machine-learning model [34]. We therefore use it as a Design Material [12] in our analysis to derive guidelines for future developers and designers of microgestures.

In contrast, black-box models are too complex to allow a straightforward analysis of their learned decision-making. Interpreting such models is an active and open research question [1, 18, 25].

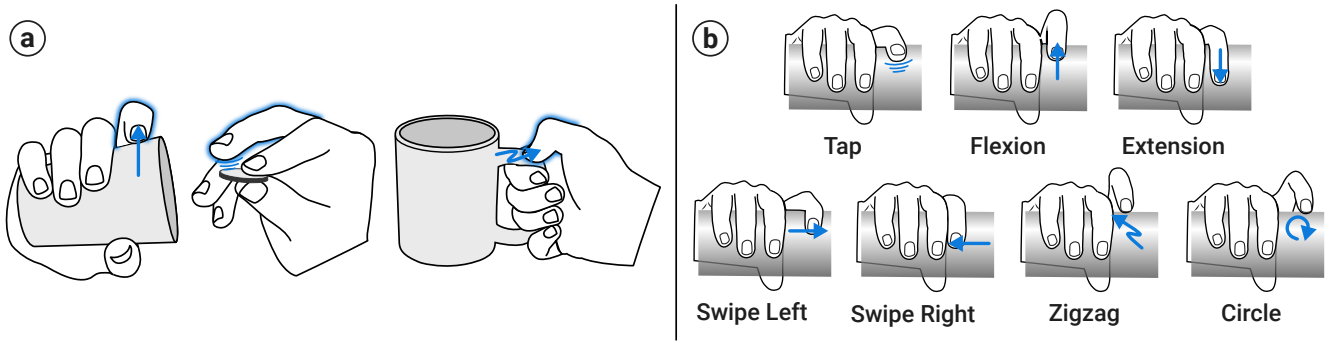


Figure 2: (a) SoloFinger microgestures are performed with a single finger on an object, while holding it. (b) Tapping or different directional movements define unique gestures that can be performed either with thumb, index, or middle finger.

Black-box models can, however, use their additional complexity to learn more advanced decision-making processes. This often results in better evaluation performance and is, therefore, usually preferred to achieve state-of-the-art results and for real-world deployment [73]. We train and evaluate such a black-box model to show a proof-of-concept implementation of our concept.

3 SOLOFINGER GESTURES

The sophistication of the human hand allows for dexterous hand-object interactions. Our fingers hold objects using a wide variety of grasps, and while manipulating objects our fingers move in versatile ways and diverse configurations. Due to this impressive richness of movement patterns, it is challenging to define unique gestures applicable across different grasp types and object geometries, yet mutually exclusive of everyday actions. We introduce the concept of SoloFinger microgestures that aim to stand out from everyday hand motions, hence reducing the likelihood of false activations.

SoloFinger microgestures are conceptually based on the observation that during everyday hand-object interaction, multiple fingers tend to move concurrently, whereas it is rare that a single finger moves extensively on the object while all others stay idle. This observation was informed by findings that finger movements tend to be highly correlated during object manipulation [20, 70]. Our work leverages this phenomenon. We ground our findings on objects that do not contain movable elements, such as mechanical buttons or sliders.

A SoloFinger gesture (Figure 2b) involves *moving a single finger by a considerable yet comfortable extent*, while *all other fingers remain static*. It is performed *while holding an object*, *with the same hand*, and *on the object* itself. SoloFinger gestures are not limited to any specific finger. We recommend using the thumb, index, or middle finger, as the ring and pinky fingers were shown to be less robust and also subjectively less preferable.

This generic approach allows for defining diverse microgestures. For instance, these comprise tapping, moving a finger forward, backward or sideways, or moving in advanced patterns, such as drawing a circle. Figure 2 (b) depicts the gestures we used in our studies. Performed with either thumb, index, or middle finger, this leads to a total of 21 gesture options we have investigated. However, more SoloFinger gestures can be conceived.

In the following, we will conceptually and practically validate the feasibility of our proposed concept. Our conceptual analysis is based on two datasets that we present in the next section. It validates two key assumptions that underlie SoloFinger gestures: extensive single-finger movements are unlikely to happen during everyday hand-object actions, and SoloFinger gestures are compatible with holding diverse types of objects in diverse grasps. Next, using a simple white-box classifier, we investigate the principled feasibility of gesture classification and show that SoloFinger gestures create little false activation during diverse everyday actions. Finally, we demonstrate the practical feasibility by presenting a proof-of-concept implementation that uses commodity hardware.

4 DATASETS FOR DAILY HAND-OBJECT ACTIONS AND SOLOFINGER

We use a data-driven method to systematically validate our concept using two datasets: a baseline dataset offering extensive coverage of everyday hand-object interaction, and a dataset of SoloFinger gestures that end users performed naturally using a wide set of grasps and actions.

4.1 Dataset with Daily Hand-Object Actions

We base our analysis of everyday object manipulation on a baseline dataset made available by prior research. Garcia-Hernando et al. [17] created the first benchmark dataset that provides precise information about hand joint positions and angles during an extensive range of hand-object interaction. It comprises data from a diverse set of 45 everyday object manipulation actions, performed with 26 objects. Data were captured using high-frequency magnetic sensors to avoid any obstruction between finger contact and object surface. Information about hand joints and fingertips was then derived using inverse kinematics. The dataset contains 105,459 RGB-D frames with 3D location of each of the 21 joints of a hand model.

We use this dataset to verify our assumption that single-finger movements are rare while grasping an object and for assessing false activations caused by SoloFinger gestures. Since our approach only focuses on handheld objects, we removed a subset of the actions from this dataset that did not involve grasping an object (performing a high-five; shaking hands; pressing the buttons of a calculator; closing liquid soap). We also combined actions with very similar finger motions and grasp types opening/closing juice and milk bottle; opening/closing peanut butter; and scratching/washing



Figure 3: The 36 actions in our dataset of SoloFinger gestures cover diverse real-world objects and grasps.

a sponge). We used the video data provided along with the dataset to inform these decisions. As shown in Fig. 3, after removal and consolidation, we are left with 36 different hand-object actions (95,788 frames). To use terminology consistent with our second dataset, we use the term "trial" to refer to the sequence of data recorded while one participant performs one action.

4.2 SoloFinger Dataset

Thus far, no studies have reported detailed hand data for single-finger movements on diverse grasp types. We therefore recorded a new dataset with hand movement data from study participants who performed SoloFinger gestures while grasping objects. Our focus here is to systematically investigate single-finger movements on a wide variety of objects. However, to evaluate gesture recognition using a more sophisticated model with multiple gesture trials and variations, we collected another dataset as described in Section 7.1.

4.2.1 SoloFinger Gestures. We centered this first study on three most basic SoloFinger gestures. These comprise the primitive finger movements of *Flexion* and *Extension*, as commonly used in the field of biomechanics, and *Tap* as a discrete motion. These 3 primitive movements are illustrated in Fig. 2 (b, top row). For a baseline comparison, we also collected one trial of static *Hold* for every case in which the object was held in a static pose and no gesture performed. Before starting the experiment, we demonstrated these 3 gestures on a cylindrical prop (which was not part of the set of objects used in the study) to familiarize participants with SoloFinger gestures.

4.2.2 Actions. We used the same 36 hand actions as in the Hand-Object Actions dataset (as shown in Fig. 3). We asked the participants to perform gestures while holding the object in a static pose.

4.2.3 Participants. We recruited 15 participants. Not all trials were recorded from 2 participants due to technical reasons. Therefore, all subsequent analysis uses data from the remaining 13 right-handed participants (6 female) aged from 21 to 30 (median=27) from different professional backgrounds (engineering, law, literature). Participants received a compensation of €20 for their participation. Before collecting data, we manually measured the hand dimensions of participants following the BigHand2.2M approach [75], which involves measuring the distance between different finger joints. We found, on average, distances from the wrist to the tips of: *thumb* - 121mm (SD=8mm), *index* - 141mm (8mm), *middle* - 151mm (9mm), *ring* - 144mm (9mm), *pinky* - 120mm (10mm).

4.2.4 Apparatus. We used the OptiTrack™ motion-tracking system with 11 cameras running at a 60Hz refresh rate to capture finger movements. We attached 8 facial reflective markers (4 mm diameter): one on each finger tip and on the wrist, and two on the MCP joint (where the finger connects to the hand) of the index and pinky fingers to help with the manual labeling. The setup is shown in Figure 4. To ensure each marker is consistently labeled with a unique ID, we manually annotated markers during post-processing in the OptiTrack's Motive software [45]. In total this results in 879,908 frames of data that we use in our analysis.

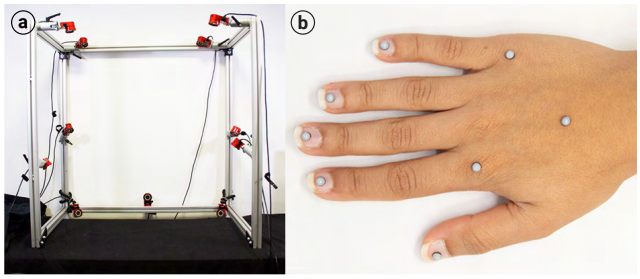


Figure 4: (a) Study setup using an OptiTrack motion capture system consisting of 11 infrared cameras. (b) Retroreflective markers placed on the hand to track finger movement.

4.2.5 Task and Procedure. The participant held an object in the dominant hand and used the same hand to perform a SoloFinger gesture on the object with a given finger. We asked participants to perform the gesture in such a way that it felt *comfortable* to them, and not in an exaggerated manner. For each action, the participants had to perform all gestures using all fingers. If they could not perform the gesture due to the odds of dropping the object, they notified the experimenter, who marked this pair of gesture-action as impossible to perform. For possible gestures, they rated each trial for ease on a 5-point Likert scale. We randomized the order of action and counterbalanced the order of gesture and finger used to perform the gesture. For each participant, the experiment took approximately 3 hours and was conducted in two sessions of 1.5 hours each. The full dataset containing recorded data for possible gestures from 13 participants with 5,530 trials is made available to the research community (see Section 1 for the link).

4.3 Data Preprocessing

We solely consider the fingertip position to assess finger movements and define the wrist position as the origin of the coordinate system. We apply a median filter on the realigned finger coordinates to reduce the noise. We analyze the hand data using overlapping windows of 1-second duration with a one-frame shift. We chose one second based on the observation that most movements were completed during this interval.

On each window, and for each finger, we create a minimal 3D bounding box to attenuate the jitter in the signal. The bounding box covers the fingertip's 3D positions during the frames that constitute this window. We calculate the diagonal of the bounding box over each window as an estimated measurement of the longest straight line the finger has moved during this window. In the following, we refer to these diagonals as the *movements* of the fingers.

5 CONCEPT VALIDATION

The two datasets described above provide data of hand-object interaction, with and without gestures. In the following, we compare them to validate our concept by assessing whether extensive single-finger movements are unlikely to happen during everyday actions. Subsequently, we evaluate the feasibility of performing SoloFinger gestures while grasping objects and derive a set of fingers we recommend for SoloFinger gestures.

5.1 SoloFinger Gestures Are Unlikely to Happen During Everyday Hand-Object Actions

The goal of this analysis is to compare finger motions during SoloFinger gestures with everyday hand-object actions. We introduce a *Peak Score* metric that quantifies how strongly the movement of a single finger deviates from the movement of the other fingers. To calculate this, we take the finger with the maximum movement and calculate the ratio between its movement and the movement of the others:

$$PeakScore = \frac{\max\limits_{f \in F} m_f}{\sum\limits_{f' \in F} m_{f'} - \max\limits_{f \in F} m_f} \quad (1)$$

where F is the set of fingers and m_f is the movement value over one window for a finger f. We compute this value for all windows in the datasets. A value of 1 shows that a single finger moved as much as the other fingers combined, whereas a value of 0.25 shows all fingers moved the exact same amount. Hence, a high score signifies that one finger traveled a considerably longer distance than the others.

The peak score allows us to numerically compare recorded actions that involve gestures with actions not involving gestures. This provides insights into whether the SoloFinger hypothesis holds. We depict the average peak scores of both datasets for each action in Figure 5. One can notice that peak scores for everyday actions are impressively low compared to actions including SoloFinger gestures. A Mann-Withney U test comparing both datasets yields a highly significant difference ($p < 0.001$ with Cohen's d=0.99). We observe an average peak score of daily hand-object actions of 0.33 (SD=0.07). Note, these actions include different finger movements from precise (like plugging the charging cable into the cellphone) to dynamic motions, such as opening or closing peanut butter, or squeezing paper. In contrast, we observe an average peak score of SoloFinger gestures of 1.70 (SD=1.95). The high standard deviations can be explained by the fact that a gesture happens quickly, hence only raising the peak score for an instant.

In addition, we noticed that the peak score for gestures depends on the grasp type and available surface area for finger movements. For example, the Tap gesture has similar peak scores on actions involving similar grasp types (e.g., `give coin` 1.56 (SD=0.90) and `tear paper` 1.49 (SD=0.70)). In contrast, other actions like `pour wine` have a different grasp type, involving all fingers in contact with the object. This provides stability to do extended gestures, resulting in higher peak scores (e.g., `pour wine` 4.19 (SD=3.55) during Tap). With respect to available surface area, the actions with smaller available surface area have smaller peak scores for Flexion and Extension when compared to Tap. This is because the object allows less room for continuous finger movement (e.g., `open peanut butter` shows a peak score of 2.99 (SD=2.66) for Tap; 1.13 (SD=0.98) and 1.08 (SD=0.71) for Flexion and Extension respectively). Further studies should investigate how other object properties, such as texture and weight affect finger movement..

Overall, these findings suggest that extensive single-finger movements are indeed rare during everyday actions, hence creating an opportunity for simple yet robust microgestures.

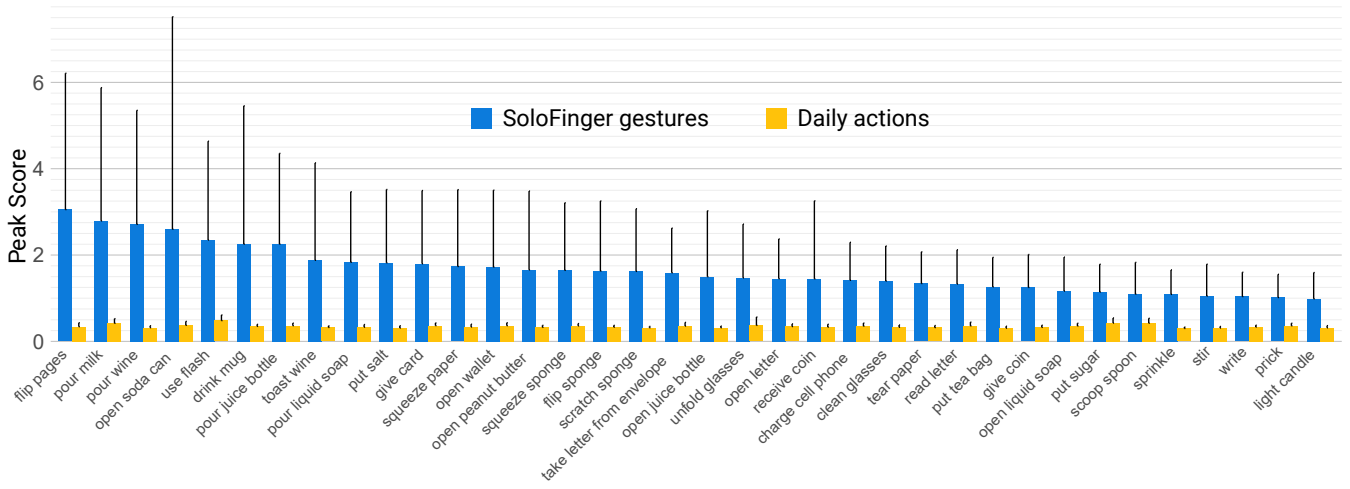


Figure 5: Average peak scores for each action present in the two datasets (i.e., with and without SoloFinger gestures). The half error bars depict one standard deviation.

5.2 SoloFinger Gestures Are Compatible with Holding Objects

We now focus on the feasibility of performing SoloFinger gestures. While holding an object, the fingers' primary task is to stabilize the object. It needs to be investigated if, despite this primary task, fingers can perform SoloFinger gestures, and if this holds true for diverse object geometries and grasps. To address these questions, we analyzed the gestures participants made in the SoloFinger dataset as well as their subjective ratings.

5.2.1 Finger individuation while holding objects. A first prerequisite for performing a SoloFinger gesture is that a single finger can move independently from others while holding the object. The *Individuation Index* (I_{ID}) [54] is a widely used metric from neuroscience that measures the extent to which a finger can move independently from others. If a finger has absolute independence, its I_{ID} is 1.00. Conversely, a value of 0.0 denotes high dependence. We calculate this metric over three windows and retain the maximum value for each trial. Results revealed a high average individuation index(>0.90) [61] for all fingers: thumb = 0.98 (SD = 0.06), index = 0.97 (0.07), middle = 0.96 (0.07), ring = 0.95 (0.07), and pinky = 0.96 (0.09). This denotes the principled possibility of performing single finger movements with all the fingers across a diverse set of actions, comprising diverse object geometries and grasps. To investigate if some actions are more suitable than others, we analysed the I_{ID} for all actions. We found that the I_{ID} is high for all actions, the lowest value being 0.93 (SD = 0.10) for the action `prick`.

Despite this principled feasibility of single-finger movement, it is obvious that not all fingers can be moved while holding an object. Depending on the grasp, some fingers are vital for stabilizing the object; moving those would cause dropping the object. For instance, while picking up a coin, the user cannot gesture with index or middle finger, but might move any other finger. In our data collection, participants attempted to execute each gesture for each action with any of the five fingers. Each combination of finger, action and gesture that a participant considered impossible to perform without dropping the object was labeled as "impossible". Noteworthy, for each action and for all participants, at least three fingers could be used to perform a SoloFinger gesture without dropping the object.

As a further metric investigating if single-finger movements are possible to perform while holding objects, we also measured the average extent of movement of performed gestures. These were: thumb 27.4 mm (SD = 21.1), index 29.1 mm (24.5), middle 36.2 mm (27.5), ring 38.3 mm (30.0), and pinky 45.7 mm (34.2). Overall, these extents indicate a sufficiently large range of motions for reliably performing gestures.

5.2.2 Ease of use. For each possible gesture, participants rated on a five-point Likert scale how easy it was for them to perform the gesture. Figure 6 shows the normalized ratings aggregated per finger. The results reveal that a vast majority of gestures performed with thumb, index and middle finger are considered easy or very easy to perform. In contrast, approximately half of the gestures performed with ring and pinky fingers were not rated as easy to perform. Mann-Whitney U tests with Bonferroni corrections revealed highly significant differences between all fingers ($p<0.001$) except between thumb-middle ($p=0.76$) and ring-pinky ($p=0.76$). This indicates users felt more comfortable performing SoloFinger gestures with the thumb, index, and middle fingers.

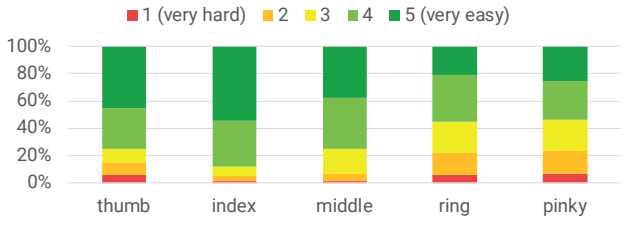


Figure 6: Normalized subjective ratings of ease-of-use of SoloFinger gestures (captured for each gesture trial).

A qualitative analysis of the video recordings of gesture trials that have received a low rating revealed that during a considerable number of these actions the ring and/or pinky finger were not in contact with the object, but spread out in mid-air. This comprised actions like `write, sprinkle, clean glasses`, where objects are grasped with thumb, index, and middle fingers mostly. This made it more difficult to perform the gesture, since the participant first had to move the finger in mid-air to bring it onto the object and then move it on the object. Several participants commented about fatigue created by the single-finger movement in mid-air, an effect that is also mentioned in prior work [61].

5.2.3 Recommended set of fingers. We conclude that single-finger microgestures can be performed with all fingers. While holding any object comprised in our dataset, a minimum of three fingers is free to move and to perform SoloFinger gestures. This shows that SoloFinger gestures are a viable input method for diverse everyday objects. However, ring and pinky fingers were oftentimes not in contact with the object. According to subjective ratings, the ease of performing gestures with these fingers is significantly reduced. While gesturing with ring and pinky fingers may still function well for select objects, we do not recommend using these fingers in systems that involve a diverse set of objects or grasps. As our goal in this paper is to investigate gestures that are compatible with versatile objects, we center our following analysis on thumb, index, and middle finger.

6 RECOGNIZING SOLOFINGER GESTURES AND FALSE ACTIVATIONS

Thus far, our findings have revealed a principled difference between SoloFinger gestures and finger movement during everyday actions and have confirmed their compatibility with grasping diverse objects. We now set out to assess in more detail the conceptual feasibility of SoloFinger gestures for robust gesture detection. We first present a white-box classification technique using thresholds. This simple model set-up enables a clear human interpretation and understanding of the prediction process. We use it to gain further insights into the use of SoloFinger gestures as well as for the evaluation of false activations in a large, pre-existing dataset of daily hand-object actions. In addition, to validate our concept and overcome the limitations of the white-box classifier, we also present a black-box classifier in Section 7.2 that uses a more powerful machine learning model and supports a more complex gesture classification setting.

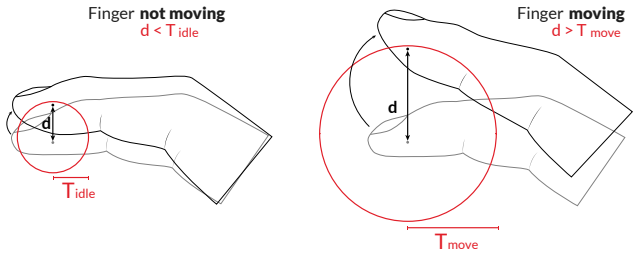


Figure 7: Two thresholds for *idle* and *moving* fingers are used to identify single-finger microgestures with the white-box classifier.

6.1 White-box Thresholding Classifier: User and Action Independent

As shown in Section 5.1, the *Peak Score* of SoloFinger gestures is much higher than in everyday hand-object actions. Given this large difference, we hypothesized that a very simple thresholding technique might be a feasible approach for gesture classification.

To understand if only the extent of finger movement contains a sufficient amount of information, we define two thresholds, one for moving and another for idle fingers, as illustrated in Figure 7. A single-finger movement is detected if the movement of a single finger is above the *move* threshold, while all others remain below the *idle* threshold.

6.1.1 Train-Test Split and Label Encoding. We used the fingertips' distance as defined in Section 4.3. For classification, we first need to identify two thresholds (idle and moving). The recordings of 3 participants were randomly selected and used to find thresholds for their data (Train set). The other 10 participants' recordings were held out and only used for evaluation (Test set). This avoids possible overfitting on evaluation data and allows us to better understand how well this procedure generalizes to unseen data. To compare the classifier's performance to the ground truth for training and evaluation, we obtained labels (gesture or non-gesture class) that were manually annotated. Each trial, consisting of a recording of a gesture or non-gesture for one specific action and finger, counts as one instance. An instance is classified as positive by the threshold model if a single-finger movement occurs in at least one of its windows. For analysis, we built a separate classifier for each gesture (Tap, Flexion or Extension). We, therefore, obtain three separate binary classifiers that check for specific gesture vs. no-gesture.

6.1.2 Threshold Optimization. During data collection, we observed that the extent of single-finger movement varies depending on the finger used and the context, i.e., how the user is grasping the hand-held object. This suggests one should define tailored thresholds for each finger and adapt these thresholds to each action. Creating individual thresholds for each action, however, is a challenging problem to solve, requiring tracking actions during user interaction, to update thresholds on-the-fly. Hence, for this conceptual evaluation, we aim for a simpler solution and define a consistent set of thresholds for all users and actions. Considering these observations, we aim at tuning 2 thresholds for our feasible set of fingers, i.e., thumb, index, middle fingers and their primitive movement, resulting in a set of $2 \times 3 \times 3$ thresholds.

We run *Bayesian Optimization* to optimize the thresholds on the training data from 3 participants [41]. We have two goals for this optimization process: 1) A gesture recognition system should only predict a gesture if the user really performed one (precision), and 2) The system should also detect as many of the user's intentional gestures as possible (recall) and not miss them. The F_1 score combines both of these goals as the harmonic mean of precision and recall. We, therefore, selected the thresholds that optimized the F_1 score on the training data. As hyperparameters for the search space for both thresholds, we used $\chi_{idle} = [0, 20]$ and $\chi_{moving} = [10, 80]$ mm. The appendix contains the optimized threshold values (see Table 2). We applied the optimized thresholds across thumb, index, and middle fingers to classify individual gestures in the Train and

Table 1: Classification performance of the white-box classifier.

	Train (3p)			Test (10p)		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Tap	1.00	0.96	0.98	1.00	0.93	0.96
Flexion	1.00	0.82	0.90	1.00	0.79	0.88
Extension	0.99	0.92	0.95	1.00	0.83	0.90

Test sets. Note that we use the same thresholds for all actions and all participants in the following analysis.

6.2 Evaluation of Gesture Recognition

Table 1 shows precision, recall, and F1 scores for the three primitive gestures, for both Train and Test datasets. On the Test dataset, a 100% precision and a recall of 93% is achieved for Tap. Flexion and Extension achieved a 100% precision and recall of 79% and 83%, respectively.

A video analysis of gesture trials with low recall revealed two main reasons for misclassification, linked to the simple distance thresholding scheme. First, several small objects offer limited surface area, resulting in smaller gestures, some of which were too small to trigger the movement threshold. For example, the `flash spray` head provides a tiny surface for fingers to slide on, resulting in an average recall of only 65% for Flexion and Extension; in contrast, Tap achieved a recall of 96% on the same surface. Second, a few actions include fingers packed closely, thereby limiting individual finger movement. Notably, `drink mug` and `pour milk` involve wrapping the fingers around a confined space, which leads to finger movement smaller than the threshold. A third source of lower recall was one action (`receive coin`) in which the object was held in the palm without any finger contact. Here, fingers were not constrained by the object, and thus idling fingers moved more extensively in mid-air, in turn violating the idle threshold.

6.3 False Activation During Daily Hand-Object Actions

For an empirical evaluation of false activation, we used the pre-existing Daily Hand-Object Actions dataset [17], which extensively covers a wide range of grasps and actions. For human interpretability, we used the white-box classification approach with idle and moving thresholds. Note, when optimizing these thresholds in Section 6.1.2, the Daily Hand-Object Actions dataset was held out. The simplicity of the white-box classifier allows us to verify on a trial-by-trial basis where the SoloFinger concept does *not* hold, i.e., under what circumstances an everyday action is misclassified as a SoloFinger gesture. We used our three sets of gesture thresholds and evaluated the whole dataset successively for each set. As the dataset does not contain any SoloFinger gesture, any detected gesture must be considered a false activation. We flagged a trial with false activation if it was triggered by any of the three gestures.

We found that false activation happened in 51 out of 933 trials. Figure 8 shows false activation scores per action. The results show that for 23 (out of 36) actions, there were no false activations. Most false activations relate to only five actions. On further analysis, we found actions with most false activations possess two main properties: the object being heavily deformable (paper) or very small (cell phone charger, match stick used to light candle). In both cases, idle fingers do not stabilize and violate the idle thresholds, because they either move along with the deformable object or move in mid-air because they are not in contact with the small object. Interestingly, the `Pour Wine` action triggered a relatively high number of false activations, despite neither a deformable nor a small object involved. When analyzing the video recordings of these trials, we noticed that four participants used a specific way of holding the bottle with thumb and index finger only, while the middle finger was suspended in air, violating the idle threshold. This was not the case with the other subjects, and as a result no false activation was triggered during their trials.

Overall, these results are encouraging and demonstrate that SoloFinger gestures, even with a very simple classification scheme, lead to little false activation during a wide range of everyday actions. They are particularly robust in cases of everyday actions that include rigid objects and involve three or more fingers in contact with the object. Most false activations related to a few specific actions. Our findings suggest that gestures performed on small objects can be more robustly classified if information about finger-object contact is available. Then, the classifier could be modified to only consider fingers while they are on the object. Classification of gestures made on deformable objects could be improved

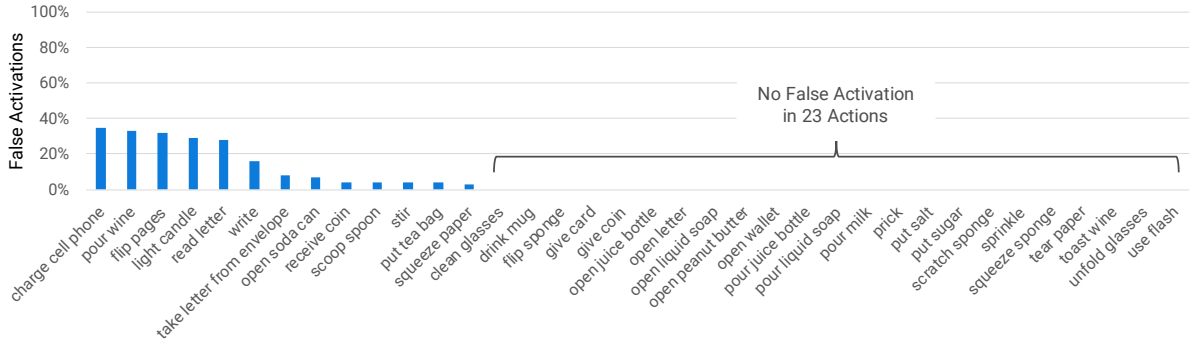


Figure 8: Occurrence of trials with false activations in the large dataset with daily hand-object actions.

with information about the position of fingertips on the object, rather than in 3D space. We will show in the next section that even without additional sensor data, classification results can be further improved by adding more feature information beyond the simple thresholds.

7 PROOF-OF-CONCEPT WITH COMMODITY HARDWARE

Our initial study confirmed the principled suitability of SoloFinger microgestures as a robust means for gestural input during diverse everyday actions. We now demonstrate a complete end-to-end recognition system with multiple trials recorded for each class. It is based on commodity hardware – a virtual reality glove – for tracking finger movements and uses a random forest classifier.

7.1 VR Glove Dataset

7.1.1 SoloFinger Gestures. In addition to the primitive finger movements investigated above (Tap, Flexion, and Extension), we added four more gesture variations: Swipe Left, Swipe Right, Zigzag, and Circle. The gestures are shown in Figure 2 (b). Each can be performed with the thumb, index or middle finger, creating a total of 21 interaction options (7 gesture variations × 3 fingers). It is worth mentioning that this is not an exclusive list, and many more variations can be created using the SoloFinger concept. In addition to gesture trials, we also recorded trials while holding the object in a static pose and while performing actions with the object.

7.1.2 Actions. To keep the study duration feasible while recording multiple gesture trials for a learning-based classifier, we selected a subset comprising actions corresponding to the five most frequently used grasps, informed by prior work [5]. (The actions are shown in Fig. 9). These actions vary considerably in their duration to complete the activity (longer grasp time), involve various motions, and possess different object geometry and rigidity.

7.1.3 Apparatus. We use the Noitom® Hi5 VR Glove to track finger movements [42]. The glove provides quaternions for each joint. To capture data in a similar format as the other datasets, we attached a cubical *Unity Game Object* at the fingertips and wrist on the provided hand model.

7.1.4 Participants. We recruited 8 right-handed participants (4 female) aged from 22 to 26 (median = 24), including two participants from the previous experiment. We used the same technique as described in our first study (see Section 4.2) to measure participants hand sizes. We found, on average, distances from the wrist to the tip of: thumb - 111mm (SD = 12mm), index - 134mm (16mm), middle - 139mm (19mm), ring - 130mm (23mm), pinky - 112mm (14mm).

7.1.5 Task and Procedure. We divided the data collection into two parts: 1) collect hand-object action data without gestures, and 2) record SoloFinger gestures. We asked half of our participants to first collect the action data and then record gesture data after a gap of approximately five days, and vice-versa with the remaining participants. For each action, the participants performed all 7 SoloFinger gestures with every possible finger of their dominant hand, except the ring finger. We recorded [8 (#participants) × (5 (#action) × 4 (#finger) × 7 (#gesture) + 5 (#hand-object action) + 5 (#static hold) - 42 (#impossible gestures))] × 10 (#trials) = 8640 trials. Informed by the empirical findings described in Section 5.2, we removed the pinky gestures and did not consider data with ring and pinky for further classification, resulting in 5,840 trials. The labeled VR Glove Data along with the precise finger movement data captured by OptiTrack are available at the link mentioned in Section 1 to facilitate future research in this area.

7.2 Black-box Classifier

The white-box classifier provided insights about the properties of SoloFinger gestures at a general level with user and action independent thresholds. The real-world deployment, however, may provide an option either to fine-tune the thresholds in a calibration process [50], or to leverage the complex decision boundaries used by the black-box classifiers to support multiple gestures. Here, we present such a system to support multiclass classification.

7.2.1 Data Preprocessing. Similar to our previous data preprocessing strategy, we used the raw 3D coordinates of the Unity Game Object and defined the wrist position as center. Subsequently, we applied a median filter on the realigned coordinates.

7.2.2 Feature Representation and Classification. TsFresh [7] is used to obtain a feature representation for each instance, i.e. for the sensor recording sequence of each individual trial. Subsequently, we fed the feature representation in a random forest classifier provided by Sci-kit Learn [47]. This pipeline is not specific to our system

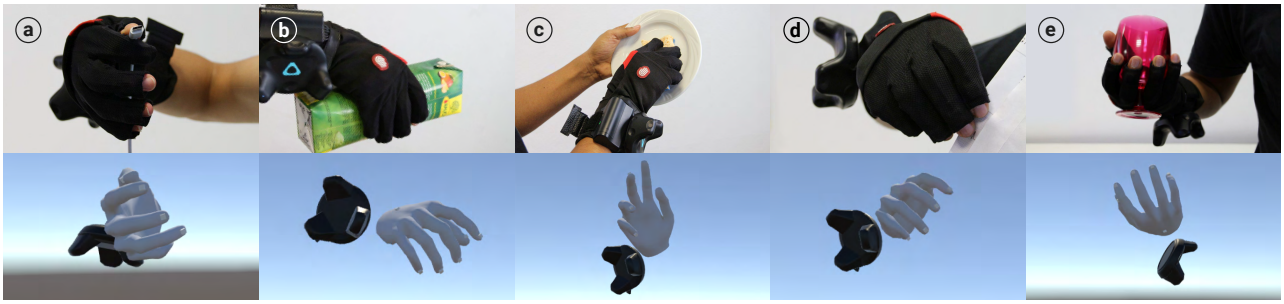


Figure 9: Proof-of-concept system using VR glove hardware supporting five frequently used grasps. (a) Charge Cell Phone, (b) Pour Juice Bottle, (c) Scratch Sponge, (d) Take Letter from Envelope, and (e) Toast Wine. The screenshots show the Unity hand model.

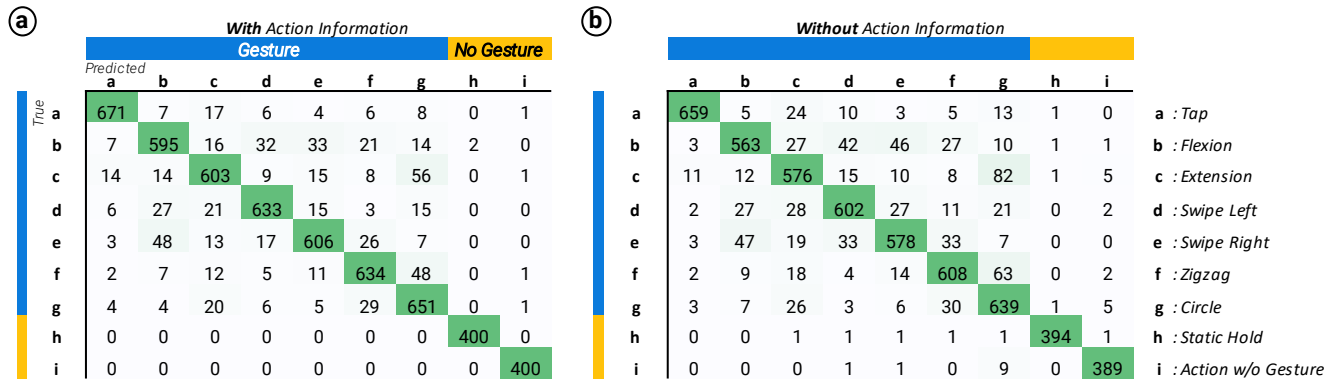


Figure 10: Confusion Matrices with and without action information.

and was used in prior work [19]. We did not optimize the hyper-parameters of the classifier and used the default settings with n-estimators = 500. Due to the personalized patterns involved in some gestures, such as Zigzag or Circle, that have a high degree of variation among users, we opted for user-dependent models. In a real setting, pre-trained models for individuals can be easily saved and restored to avoid the burden of per-session training. From our initial analysis, we learned that finger movements vary on different objects. Therefore, we performed evaluation in two conditions - training with and without action information. Thus, the classification task was to classify individual trials of all 9 classes (7 gestures + 1 static hold + 1 action without gesture) without action information for every participant. For the condition with action information, we trained models separately for all five actions.

We evaluate classification using the leave-one-trial-out 10-fold cross validation technique for each participant (9 Train trials of each class, 1 Test trial - 10 permutations). Note that the imbalance in the trial count is because we combine the results across all actions which include impossible gestures on a few fingers.

7.3 Results

7.3.1 With Action Information. Figure 10 (a) shows the result of the gesture classification for a setting in which the action is known. Note that no false activations were triggered. The average accuracy is 89%. Across all actions, the Tap gesture achieved the highest accuracy of 93%, followed by Circle with 90%. The lowest accuracy was 83% for Flexion and 84% for Extension. We assume this is related to the VR glove's wiring, which runs across the back of the fingers and might have restricted bending movements. In contrast to results from the white-box classifier, where fingers violated the idle threshold on extremely deformable objects (paper), the black-box classification did not trigger any false activations for the action that involved paper (`take letter from envelope`).

Note that this model requires prior knowledge about the action the user is performing. In many applications, this information is readily available. For instance, a microgestural input system devised for a specific object (such as a smart surgical tool, an augmented drilling machine, or a smart pen) can readily use an object-specific model. Otherwise, activity recognition [68] could be integrated to identify the ongoing action.

7.3.2 Without Action Information. For comparison, we report results for the more demanding case in which the action is not known. Here, the average classification accuracy is 86%. Here again Tap achieved the highest accuracy of 92%. The lowest accuracy of 84% was Circle, which sometimes gets misclassified with Zigzag, particularly when performed with the middle finger. False activations were triggered only in 2.12% of total action and static hold trials, with a total of 17 (out of 800) trials (see Figure 10 (b)).

8 DISCUSSION, LIMITATIONS AND FUTURE WORK

Our findings demonstrate that SoloFinger gestures provide a robust scheme for microgestural input on the object itself, creating a low number of false activations during many everyday hand-object actions. Here we reflect on the strength and limitations of the proposed approach and discuss areas for future work.

8.1 Sensing Technology to Implement SoloFinger

SoloFinger is not restricted to the hardware we use in our proof-of-concept implementation and can be implemented with various sensing technologies. A sensor should provide information about the distance that a user's fingertips move, at approximately a 2–3 mm resolution. Of note, the sensor does not necessarily need to provide information about finger-object contact. However, an important requirement is that it function while the fingertip is in contact with an object. While this prevents us from using established computer vision techniques, which tend to suffer from occlusion generated by the object [62], recent advances show promising results for hand-object interaction [39, 62, 66]. Objects equipped with high-resolution touch sensors [4, 49, 60, 64] are also promising to further deploy SoloFinger gestures, given the precise temporal and spatial contact information provided. Furthermore, other technical approaches such as magnetic [6, 24], electro-magnetic [46], radar-based [33], and IMU-based approaches [59] can be promising avenues for realizing SoloFinger gestures.

8.2 Gesture Classification

Our results revealed that false activations are primarily caused by deformable and small objects, such as paper or a match stick. Our current scheme considers data from fingers no matter whether in contact with the object or not. In our white-box analysis, we found that in-air fingers tend to make considerably larger involuntary movements while another finger is gesturing, due to the lack of stabilizing object contact. Hence in-air fingers more frequently violated the idle threshold and therefore led to lower recall. Future implementations could detect finger contact using a dedicated sensor or approximate it based on grasp or action type, and then only consider fingers that are in contact with the object for classification. We showed that adding more features and a more advanced classifier as in our black-box implementation can further help increase classification performance and robustness. Furthermore, using an ensemble of classifiers may further improve the accuracy of gesture detection: a binary classifier, as described in the white-box classification, forms a first layer to identify the gesture and non-gesture class; this is followed by a second layer of multi-class classification.

The SoloFinger gestures in our study were performed with the object held in a static pose. Future work should investigate the effect of performing gestures while manipulating the object, such as hammering with a hammer. Note, however, that our analysis of false activation *did* involve object manipulation. Also, we used one set of thresholds for all users; it will need to be investigated whether this generalizes to children or users with very large hands. Future work could normalize thresholds for hand size or finger length. We intentionally focused on a basic set of gestures for this first study. In future work, we plan to develop a design tool, inspired by [28], that offers computational assistance to designers seeking to identify new SoloFinger gestures.

8.3 Investigating More Objects and Specialized Actions

While our analysis did demonstrate that SoloFinger gestures could be very robust in reducing false activations during daily hand-object interactions using diverse objects, there remain additional cases to be investigated. First, our study did not include objects that comprise mechanical interface elements, such as buttons or sliders, or touchscreens. We assume that using those might involve extensive single-finger movements similar to SoloFinger gestures. In these cases, we recommend that the designer should carefully understand the regular finger movement on such objects and then select SoloFinger gestures accordingly to avoid conflicts. Second, we observed that highly deformable objects can result in increased false activation. Our datasets do not contain information about the touched location on an object; hence it does not allow us to differentiate between finger movements on an object and fingers that remain at the same object location but move in 3D space while the object itself is deforming. With a sensing technology that provides on-object touch location, we anticipate that the SoloFinger concept could still work for most deformable objects. Furthermore, the effect of specialized dexterous actions need to be studied. Actions such as playing a musical instrument, sculpting or performing a surgery might possibly involve more pronounced single-finger movements.

8.4 Impact of the Dataset

While we have tested false positives on an extensive dataset covering a large set of grasps, objects, and actions, the dataset is subject to limitations. Our motive behind using this dataset rather than a field recording is that it offers precise and realistic hand data with broad coverage of hand-object actions. Given the technical limitations of recording such highly articulate actions, this would likely be challenging in a field recording. Specifically, the dataset contains focused activities, which we worked on as a first step to explore our concept's potential. However, in addition to the focused activities, different idle phases may occur in real-world interaction. Idle phases might bring additional challenges for classification, for instance arising during change of hands, multi-tasking, nervous tapping, or when the user is fiddling with an object. These should be investigated in future work by capturing longer-term in-the-wild data.

9 CONCLUSION

We presented *SoloFinger*, a novel concept to identify and design robust single-finger microgestures while grasping everyday objects. The results from our data-driven analysis confirmed the insight that fingers tend to be either static or moving concurrently while holding and manipulating a wide range of objects. This opens up a space for rapid, easy and elegant microgestures performed on the object itself and resilient to false activations. Our simple white-box classifier achieved an average precision of 100% and recall of 88%, with only 51 false activations among 933 action trials of an unseen dataset. Of note, no false activation was triggered in 23 actions out of 36 actions.

We ultimately presented a proof-of-concept with commodity hardware and a black-box classifer that can detect 7 types of SoloFinger microgestures and hand actions with an accuracy of 89%. When the action is known, no false activations occurred in the collected dataset of around 800 everyday actions, whereas a small number (2.12%) of trials involved false activation in the more demanding case when a single classifier is used for all actions.

During this work, we tested several state-of-the art techniques to capture as much information as possible during hand-object manipulation, but these techniques still suffer from occlusion. Inspired by our peers [13, 30], we also provide our dataset to the community to further advance the understanding of dexterity of single-finger movements while grasping objects and leverage this dexterity to design quick and seamless gestures that can be integrated with everyday actions.

ACKNOWLEDGMENTS

We thank all participants of our user studies. We are grateful to Frank Beruscha and Thorsten Sohnke from Bosch Research for their invaluable input on applications and our work's practicality. We also like to thank Jochen Huber, Marion Koelle and Paul Strohmeier for their feedback; Muhammad Hamid and Akhmad Ali Sajidin for illustrations. This work received funding from Bosch Research and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 714797).

REFERENCES

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM. https://doi.org/10.1145/3290605.3300233

[2] Fraser Anderson, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2015. Supporting Subtlety with Deceptive Devices and Illusory Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*. ACM. https://doi.org/10.1145/2702123.2702336

[3] Leonardo Angelini, Francesco Carrino, Stefano Carrino, Maurizio Caon, Omar Abou Khaled, Jürgen Baumgartner, Andreas Sonderegger, Denis Lalanne, and Elena Mugellini. 2014. Gesturing on the Steering Wheel. In *Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '14)*. ACM. https://doi.org/10.1145/2667317.2667414

[4] Kazuyuki Arimatsu and Hideki Mori. 2020. Evaluation of Machine Learning Techniques for Hand Pose Estimation on Handheld Device with Proximity Sensor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM. https://doi.org/10.1145/3313831.3376712

[5] Ian M. Bullock, Joshua Z. Zheng, Sara De La Rosa, Charlotte Guertler, and Aaron M. Dollar. 2013. Grasp Frequency and Usage in Daily Household and Machine Shop Tasks. *IEEE Transactions on Haptics*. https://doi.org/10.1109/toh.2013.6

[6] Liwei Chan, Rong-Hao Liang, Ming-Chang Tsai, Kai-Yin Cheng, Chao-Huai Su, Mike Y. Chen, Wen-Huang Cheng, and Bing-Yu Chen. 2013. FingerPad: private and subtle interaction using fingertips. In *Proceedings of the SIGCHI Symposium on User Interface Software and Technology (UIST '13)*. ACM. https://doi.org/10.1145/2501988.2502016

[7] Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W. Kempa-Liehr. 2018. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing* 307. https://doi.org/10.1016/j.neucom.2018.03.067

[8] Vanessa Cobus and Wilko Heuten. 2019. To Beep or Not to Beep? Evaluating Modalities for Multimodal ICU Alarms. *Multimodal Technologies and Interaction* 3. https://doi.org/10.3390/mti3010015

[9] M.R. Cutkosky. 1989. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on Robotics and Automation* 5. https://doi.org/10.1109/70.34763

[10] Travis Deyle, Szabolcs Palinko, Erika Shehan Poole, and Thad Starner. 2007. Hambone: A Bio-Acoustic Gesture Interface. In *IEEE International Symposium on Wearable Computers (ISWC '07)*. IEEE. https://doi.org/10.1109/iswc.2007.4373768

[11] Tanja Döring, Dagmar Kern, Paul Marshall, Max Pfeiffer, Johannes Schöning, Volker Gruhn, and Albrecht Schmidt. 2011. Gestural interaction on the steering wheel. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems (CHI '11)*. ACM. https://doi.org/10.1145/1978942.1979010

[12] Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX Design Innovation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM. https://doi.org/10.1145/3025453.3025739

[13] Anna Maria Feit, Daryl Weir, and Antti Oulasvirta. 2016. How We Type: Movement Strategies and Performance in Everyday Typing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM. https://doi.org/10.1145/2858036.2858233

[14] Thomas Feix, Javier Romero, Heinz-Bodo Schmiedmayer, Aaron M. Dollar, and Danica Kragic. 2016. The GRASP Taxonomy of Human Grasp Types. *IEEE Transactions on Human-Machine Systems*. https://doi.org/10.1109/thms.2015.2470657

[15] George W. Fitzmaurice, Hiroshi Ishii, and William A. S. Buxton. 1995. Bricks: laying the foundations for graspable user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*. ACM. https://doi.org/10.1145/223904.223964

[16] Euan Freeman, Gareth Griffiths, and Stephen A. Brewster. 2017. Rhythmic micro-gestures: discreet interaction on-the-go. In *Proceedings of the International Conference on Multimodal Interaction (ICMI '17)*. ACM. https://doi.org/10.1145/3136755.3136815

[17] Guillermo Garcia-Hernando, Shanxin Yuan, Seungryul Baek, and Tae-Kyun Kim. 2018. First-Person Hand Action Benchmark with RGB-D Videos and 3D Hand Pose Annotations. In *Proceedings of Computer Vision and Pattern Recognition (CVPR '18)*. IEEE. https://doi.org/10.1109/cvpr.2018.00050

[18] Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. 2018. Explaining Explanations: An Overview of Interpretability of Machine Learning. In *International Conference on Data Science and Advanced Analytics (DSAAI '18)*. IEEE. https://doi.org/10.1109/dsaa.2018.00018

[19] Jun Gong, Yang Zhang, Xia Zhou, and Xing-Dong Yang. 2017. Pyro: Thumb-Tip Gesture Recognition Using Pyroelectric Infrared Sensing. In *Proceedings of the SIGCHI Symposium on User Interface Software and Technology (UIST '17)*. ACM. https://doi.org/10.1145/3126594.3126615

[20] Victor Gonzalez-Sanchez, Jennifer Rowson, and Alaster Yoxall. 2016. Analysis of finger movement coordination during the Variable Dexterity Test and comparative activities of daily living. *International Journal of Therapy and Rehabilitation* 23. https://doi.org/10.12968/ijtr.2016.23.10.481

[21] Daniel Groeger and Jürgen Steimle. 2018. ObjectSkin: Augmenting Everyday Objects with Hydroprinted Touch Sensors and Displays. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT '18)*. ACM. https://doi.org/10.1145/3161165

[22] Seongkook Heo, Michelle Annett, Benjamin Lafreniere, Tovi Grossman, and George Fitzmaurice. 2017. No Need to Stop What You'Re Doing: Exploring No-Handed Smartwatch Interaction. In *Proceedings of the Graphics Interface Conference (GI '17)*. Canadian Human-Computer Communications Society. https://dl.acm.org/doi/10.5555/3141475.3141498

[23] Ken Hinckley, Patrick Baudisch, Gonzalo Ramos, and Francois Guimbretiere. 2005. Design and analysis of delimiters for selection-action pen gesture phrases in scriboli. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM. https://doi.org/10.1145/1054972.1055035

[24] Da-Yuan Huang, Liwei Chan, Shuo Yang, Fan Wang, Rong-Hao Liang, De-Nian Yang, Yi-Ping Hung, and Bing-Yu Chen. 2016. DigitSpace: Designing Thumb-to-Fingers Touch Interfaces for One-Handed and Eyes-Free Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM. https://doi.org/10.1145/2858036.2858483

[25] Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. 2020. Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM. https://doi.org/10.1145/3313831.3376219

[26] Ryo Kawahata, Atsushi Shimada, Takayoshi Yamashita, Hideaki Uchiyama, and Rin ichiro Taniguchi. 2016. Design of a Low-false-positive Gesture for a Wearable Device. In *Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM '16)*. SCITEPRESS. https://doi.org/10.5220/0005701905810588

[27] Frederic Kerber, Philipp Schardt, and Markus Löchtefeld. 2015. WristRotate: a personalized motion gesture delimiter for wrist-worn devices. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia (MUM '15)*. ACM. https://doi.org/10.1145/2836041.2836063

[28] Daniel Kohlsdorf, Thad Starner, and Daniel Ashbrook. 2011. MAGIC 2.0: A web tool for false positive prediction and prevention for gesture recognition systems. In *International Conference on Automatic Face and Gesture Recognition 2011*. IEEE. https://doi.org/10.1109/fg.2011.5771412

[29] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In *Proceedings of the SIGCHI Symposium on User Interface Software and Technology (UIST '16)*. ACM. https://doi.org/10.1145/2984511.2984582

[30] Huy Viet Le, Sven Mayer, and Niels Henze. 2018. InfiniTouch: Finger-Aware Interaction on Fully Touch Sensitive Smartphones. In *Proceedings of the SIGCHI Symposium on User Interface Software and Technology (UIST '18)*. ACM. https://doi.org/10.1145/3242587.3242605

[31] Huy Viet Le, Sven Mayer, Benedict Steuerlein, and Niels Henze. 2019. Investigating Unintended Inputs for One-Handed Touch Interaction Beyond the Touchscreen. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '19)*. ACM. https://doi.org/10.1145/3338286.3340145

[32] Hanchuan Li, Can Ye, and Alanson P. Sample. 2015. IDSense: A human object interaction detection system based on passive UHF RFID. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*. ACM. https://doi.org/10.1145/2702123.2702178

[33] Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: ubiquitous gesture sensing with millimeter wave radar. In *Transactions on Graphics (SIGGRAPH '16)*. ACM. https://doi.org/10.1145/2897824.2925953

[34] Zachary C. Lipton. 2018. The mythos of model interpretability. In *Communications of the ACM*. ACM. https://doi.org/10.1145/3233231

[35] Hao Lu and Yang Li. 2015. Gesture On: Enabling Always-On Touch Gestures for Fast Mobile Access from the Device Standby Mode. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*. ACM. https://doi.org/10.1145/2702123.2702610

[36] Christine L MacKenzie and Thea Iberall. 1994. *The Grasping Hand, Volume 104*. Elsevier.

[37] Fabrice Matulic, Riku Arakawa, Brian Vogel, and Daniel Vogel. 2020. PenSight: Enhanced Interaction with a Pen-Top Camera. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM. https://doi.org/10.1145/3313831.3376147

[38] Fabrice Matulic, Brian Vogel, Naoki Kimura, and Daniel Vogel. 2019. Eliciting Pen-Holding Postures for General Input with Suitability for EMG Armband Detection. In *Proceedings of the ACM International Conference on Interactive Surfaces and Spaces (ISS '19)*. ACM. https://doi.org/10.1145/3343055.3359720

[39] Franziska Mueller, Florian Bernard, Oleksandr Sotnychenko, Dushyant Mehta, Srinath Sridhar, Dan Casas, and Christian Theobalt. 2018. GANerated Hands for Real-Time 3D Hand Tracking from Monocular RGB. In *Proceedings of Computer Vision and Pattern Recognition (CVPR '18)*. IEEE. https://handtracker.mpi-inf.

mpg.de/projects/GANeratedHands/
[40] John Russell. Napier and Russell H. Tuttle. 1993. *Hands*. Princeton University Press.
[41] Fernando Nogueira. 2014–. Bayesian Optimization: Open source constrained global optimization tool for Python. https://github.com/fmfn/BayesianOptimization
[42] Noitom. 2020 (accessed August 10, 2020). Hi5 VR GLOVE. https://hi5vrglove.com
[43] Donald A. Norman. 1998. *The design of everyday things*. MIT.
[44] 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 SIGCHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM. https://doi.org/10.1145/3313831.3376236
[45] OptiTrack. 2020 (accessed August 10, 2020). Motive - Optical motion capture software. https://optitrack.com/products/motive/
[46] Farshid Salemi Parizi, Eric Whitmire, and Shwetak Patel. 2019. AuraRing: Precise Electromagnetic Finger Tracking. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT '19)*, Vol. 3. ACM. https://doi.org/10.1145/3369831
[47] 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).
[48] Patryk Pomykalski, Mikołaj P. Woźniak, Paweł W. Woźniak, Krzysztof Grudzień, Shengdong Zhao, and Andrzej Romanowski. 2020. Considering Wake Gestures for Smart Assistant Use. In *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems (CHI EA '20)*. ACM. https://doi.org/10.1145/3334480.3383089
[49] Narjes Pourjafarian, Anusha Withana, Joseph A. Paradiso, and Jürgen Steimle. 2019. Multi-Touch Kit: A Do-It-Yourself Technique for Capacitive Multi-Touch Sensing Using a Commodity Microcontroller. In *Proceedings of the SIGCHI Symposium on User Interface Software and Technology (UIST '19)*. ACM. https://doi.org/10.1145/3332165.3347895
[50] Philip Quinn, Seungyon Claire Lee, Melissa Barnhart, and Shumin Zhai. 2019. Active Edge: Designing Squeeze Gestures for the Google Pixel 2. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM. https://doi.org/10.1145/3290605.3300504
[51] Christian Rendl, David Kim, Patrick Parzer, Sean Fanello, Martin Zirkl, Gregor Scheipl, Michael Haller, and Shahram Izadi. 2016. FlexCase: Enhancing Mobile Interaction with a Flexible Sensing and Display Cover. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM. https://doi.org/10.1145/2858036.2858314
[52] Jaime Ruiz and Yang Li. 2011. DoubleFlip: a motion gesture delimiter for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM. https://doi.org/10.1145/1978942.1979341
[53] 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 SIGCHI Symposium on User Interface Software and Technology (UIST '09)*. ACM. https://doi.org/10.1145/1622176.1622208
[54] M. H. Schieber. 1991. Individuated finger movements of rhesus monkeys: a means of quantifying the independence of the digits. *Journal of Neurophysiology* 65. https://doi.org/10.1152/jn.1991.65.6.1381
[55] G. Schlesinger. 1919. Der mechanische Aufbau der künstlichen Glieder. In *Ersatzglieder und Arbeitshilfen*. Springer Berlin Heidelberg, 321–661. https://doi.org/10.1007/978-3-662-33009-8_13
[56] Philipp Schoessler, Sang won Leigh, Krithika Jagannath, Patrick van Hoof, and Hiroshi Ishii. 2015. Cord UIs: Controlling Devices with Augmented Cables. In *Proceedings of the International Conference on Tangible, Embedded, and Embodied Interaction (TEI '15)*. ACM. https://doi.org/10.1145/2677199.2680601
[57] Marcos Serrano, Eric Lecolinet, and Yves Guiard. 2013. Bezel-Tap Gestures: Quick Activation of Commands from Sleep Mode on Tablets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM. https://doi.org/10.1145/2470654.2481421
[58] Adwait Sharma, Joan Sol Roo, and Jürgen Steimle. 2019. Grasping Microgestures: Eliciting Single-hand Microgestures for Handheld Objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM. https://doi.org/10.1145/3290605.3300632
[59] Yilei Shi, Haimo Zhang, Kaixing Zhao, Jiashuo Cao, Mengmeng Sun, and Suranga Nanayakkara. 2020. Ready, Steady, Touch!: Sensing Physical Contact with a Finger-Mounted IMU. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT '20)*, Vol. 4. ACM. https://doi.org/10.1145/3397309
[60] Hyunyoung Song, Hrvoje Benko, Francois Guimbretiere, Shahram Izadi, Xiang Cao, and Ken Hinckley. 2011. Grips and gestures on a multi-touch pen. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM. https://doi.org/10.1145/1978942.1979138
[61] Srinath Sridhar, Anna Maria Feit, Christian Theobalt, and Antti Oulasvirta. 2015. Investigating the Dexterity of Multi-Finger Input for Mid-Air Text Entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*. ACM. https://doi.org/10.1145/2702123.2702136
[62] Srinath Sridhar, Franziska Mueller, Antti Oulasvirta, and Christian Theobalt. 2015. Fast and Robust Hand Tracking Using Detection-Guided Optimization. In *Proceedings of Computer Vision and Pattern Recognition (CVPR '15)*. IEEE. http://handtracker.mpi-inf.mpg.de/projects/FastHandTracker/
[63] Thad Starner. 2013. Project Glass: An Extension of the Self. *IEEE Pervasive Computing* 12 (April 2013). https://doi.org/10.1109/mprv.2013.35
[64] Brandon T. Taylor and V Michael Bove. 2008. The Bar of Soap: A Grasp Recognition System Implemented in a Multi-functional Handheld Device. In *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM. https://doi.org/10.1145/1358628.1358874
[65] Brandon T. Taylor and V. Michael Bove. 2009. Graspables: grasp-recognition as a user interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM. https://doi.org/10.1145/1518701.1518842
[66] Bugra Tekin, Federica Bogo, and Marc Pollefeys. 2019. H+O: Unified Egocentric Recognition of 3D Hand-Object Poses and Interactions. In *Proceedings of Computer Vision and Pattern Recognition (CVPR '19)*. IEEE. https://doi.org/10.1109/cvpr.2019.00464
[67] Bryan Wang and Tovi Grossman. 2020. BlyncSync: Enabling Multimodal Smartwatch Gestures with Synchronous Touch and Blink. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM. https://doi.org/10.1145/3313831.3376132
[68] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. 2019. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters* 119. https://doi.org/10.1016/j.patrec.2018.02.010
[69] Daniel Wigdor and Dennis Wixon. 2011. *Brave NUI world : designing natural user interfaces for touch and gesture*. Morgan Kaufmann/Elsevier.
[70] John Williamson. 2016. Fingers of a Hand Oscillate Together: Phase Syncronisation of Tremor in Hover Touch Sensing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM. https://doi.org/10.1145/2858036.2858235
[71] Raphael Wimmer. 2010. Grasp sensing for human-computer interaction. In *Proceedings of the International Conference on Tangible Embedded and Embodied Interaction (TEI '10)*. ACM. https://doi.org/10.1145/1935701.1935745
[72] Katrin Wolf. 2016. Microgestures—Enabling Gesture Input with Busy Hands. In *Peripheral Interaction: Challenges and Opportunities for HCI in the Periphery of Attention*. Springer International Publishing. https://doi.org/10.1007/978-3-319-29523-7_5
[73] Katrin Wolf, Sven Mayer, and Stephan Meyer. 2016. Microgesture detection for remote interaction with mobile devices. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '16)*. ACM. https://doi.org/10.1145/2957265.2961865
[74] Katrin Wolf, Anja Naumann, Michael Rohs, and Jörg Müller. 2011. A Taxonomy of Microinteractions: Defining Microgestures Based on Ergonomic and Scenario-Dependent Requirements. In *Human-Computer Interaction – INTERACT 2011*. Springer. https://doi.org/10.1007/978-3-642-23774-4_45
[75] Shanxin Yuan, Qi Ye, Bjorn Stenger, Siddhant Jain, and Tae-Kyun Kim. 2017. BigHand2.2M Benchmark: Hand Pose Dataset and State of the Art Analysis. In *Proceedings of Computer Vision and Pattern Recognition (CVPR '17)*. IEEE. https://doi.org/10.1109/cvpr.2017.279
[76] Yang Zhang, Gierad Laput, and Chris Harrison. 2017. Electrick: Low-Cost Touch Sensing Using Electric Field Tomography. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM. https://doi.org/10.1145/3025453.3025842

A APPENDIX

Table 2: Optimized threshold values (in mm).

Gesture	Thumb		Index		Middle	
	Idle	**Move**	**Idle**	**Move**	**Idle**	**Move**
Tap	16.94	22.45	12.73	21.99	7.66	15.59
Flexion	8.94	20.90	9.64	19.05	6.97	15.49
Extension	10.96	21.48	10.05	16.32	9.35	13.00