Master's Thesis

Sonar Sensing on Unmodified Smartwatch

Jiwan Kim

Department of Design

Ulsan National Institute of Science and Technology

2024

Sonar Sensing on Unmodified Smartwatch

Jiwan Kim

Department of Design

Ulsan National Institute of Science and Technology

Sonar Sensing on Unmodified Smartwatch

A thesis/dissertation submitted to the
Ulsan National Institute of Science and Technology (UNIST)
in partial fulfillment of the
requirements for the degree of
Master of Science

Jiwan Kim

January 9th, 2024 of submission

Approved by

Advisor

Kyungho Lee

ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Sonar Sensing on Unmodified Smartwatch

Jiwan Kim

This certifies that the thesis/dissertation of Jiwan Kim is approved.

January 9th, 2024 of submission

signature

____

Advisor: Prof. Kyungho Lee

signature

____

Prof. Young-woo Park

signature

____

Prof. Ian Oakley

Abstract

Smartwatches are used by millions of people for applications in health, finance, and communication. However, their diminutive screen size limits the expressivity and security encompassing touchscreen interaction. To address this issue, this thesis explored two novel sonar-based approaches, each for general and user-specific perspectives.

Our first exploration targets the realm of general interaction, specifically finger identification. Despite the recognized potential of finger identification for enhancing smartwatch expressivity, its implementation remains challenging, often relying on external devices (e.g., worn magnets) or explicit instructions. Addressing these limitations, this paper explores a novel approach to natural and unencumbered finger identification on an unmodified smartwatch: sonar. To do this, we adapt an existing finger-tracking smartphone sonar implementation—rather than extract finger motion, we process raw sonar fingerprints representing the complete sonar scene recorded during a touch. We capture data from 16 participants operating a smartwatch and use their sonar fingerprints to train a deep learning recognizer that identifies taps by the thumb, index, and middle fingers with an accuracy of up to 93.7%, sufficient to support meaningful application development.

We then pivot to a user-specific angle, specifically user authentication. While various user authentication technologies have been extensively explored in smartphone use scenarios, the applicability of these approaches to smartwatches is typically limited due to the small watch form factor. To improve authentication on smartwatches, we propose SonarAuth, a novel user authentication system for unmodified commercial smartwatches using behavioral biometrics derived from motion, touch, and around-device motions. We collected data from 24 participants from single touch to the watch screen with the thumb, index, and middle fingers. Using a multimodal deep learning classifier, we achieved a promising mean Equal Error Rate(EER) of 6.41% for user authentication based on a single thumb tap. We note that our system is usable and has good potential to be combined with other authentication modalities.

Through this holistic investigation, the thesis highlights the transformative capability of sonar sensing in unmodified smartwatches, forging a path for more intuitive and secure wearable interactions in the real world.

Contents

List of Figures

List of Tables

Publications

This dissertation is based on the following peer-reviewed publications, for which I am the first author.

- **SonarID: Using Sonar to Identify Fingers on a Smartwatch**

  Kim, J., & Oakley, I. (2022, April). SonarID: Using Sonar to Identify Fingers on a Smartwatch. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1-10).

- **SonarAuth: Using Around Device Sensing to Improve Smartwatch Behavioral Biometrics**

  Kim, J., Park, J., & Oakley, I. (2023, October). SonarAuth: Using Around Device Sensing to Improve Smartwatch Behavioral Biometrics. In Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing (pp. 83-87).

In addition, I contributed to other publications during my Master degree. The following publications are not included in this thesis.

- Lee, D., Kim, J., & Oakley, I. (2021, May). Fingertext: Exploring and optimizing performance for wearable, mobile and one-handed typing. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-15).
- Kim, J., & Gil, H. (2022, October). Top-Levi: Multi-User Interactive System Using Acoustic Levitation. In Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (pp. 1-3).
- Kim, J., Lee, D., Kim, J., & Oakley, I. (2023, February). Can Eye Gaze Improve Emotional State Detection on Off the Shelf Smart Devices Jiwan Kim Doyoung Lee Jaeho Kim. In 2023 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 378-380). IEEE.

I appreciate all my collaborators who have been an integral part of my Master's journey.

I Introduction

1.1 Motivations

Smartwatches are increasingly powerful personal computers capable of a wide range of advanced functions such as tracking health status [1], physiological signals [2] and motor behaviors [3], mediating financial transactions [4], displaying messages and notifications [5], and supporting scheduling and navigation activities [6]. However, their small size (typically in the order of 3cm by 3cm) limits the expressiveness of conventional touch screen input techniques—there is limited space to present and select on-screen targets, and the fat-finger problem [7, 8] means that much of the watch screen is obscured during interaction.

A very wide body of research has sought to extend the input capabilities of smartwatches through techniques such as around-device interaction [9, 10], finger gesture sensing [11, 12], or augmenting the screen to detect additional touch properties such as pressure [13], contact area [14], the spatial [15] or temporal patterns of multiple touches [16], or the ability to distinguish between touching fingers [17]. In this latter approach, different functions have been assigned to each digit to support both general target selection tasks [18] and also the skilled performance of high bandwidth activities such as typing [19]. Authors have argued the modality is a good fit for the smartwatch form factor: it can be readily understood [18] and performed [19] and is rich enough to support a wide range of interaction techniques [17]. While the potential of finger identification is clear, practical and effective implementations remain challenging. For example, previous works relied on additional worn hardware, such as optical distance sensors mounted on nails [19] and magnetic ring [17], or explicit instructions [18].

Not only for broad input capabilities, but limited interaction on small touch screens also impacts the validity of user authentications. As various smartwatch applications inevitably require access to highly sensitive personal information and services, such as analyzing health and fitness data [20, 21], storing personal credentials [22] and making financial transactions [23, 24], it is important to secure access to smartwatches by, for example, implementing authentication techniques, such as lock screens. Indeed, this is becoming common practice. PIN and/or pattern [25], two device lock schemes popular on smartphones, have long been built into the smartwatch platforms. The use of authentication technologies is also becoming mandatory on smartwatches. However, while the principles are sound, the practice is more challenging. Users may be reluctant to enable lock screens due to the burden involved in authenticating [26, 27]. Traditional authentication input systems may directly contribute to this: entering complex information precisely on the small screen of a smartwatch can be slow, or error prone [26] due to issues such as the fat-finger problem.

1.2 Sonar sensing on smartwatches

To address these issues, this thesis explored two novel sonar sensing on an unmodified smartwatch, each for general interaction, especially finger identification, and user-specific interaction, especially user authentication. This modality requires no additional hardware and has the ability to detect the around-movement above a smartwatch, potentially supporting hover input in addition to touch. Reflecting these advantages, sonar has been previously proposed as a technique for around-device interaction on a smartwatch. FingerIO [28] presented a watch-based sonar system that tracks finger movements to support gestural or pointing input adjacent to a device. An extensive body of literature has also explored sonar on smartphones in tasks as diverse as phone grip pose identification [29], back of device input [30], breathing detection [31] and health monitoring [32]. While the smartphone literature is quite mature, the smartwatch literature is relatively limited, and we are not aware of prior work applying sonar to identify touching fingers or authenticate users in wrist-wearable contexts. We seek to fill this gap by presenting a sonar implementation for an off-the-shelf smartwatch and a study that captures the sonar fingerprints generated by smartwatch touches. In this thesis, we evaluated our system by conducting two different studies for different purposes.

First, to validate the sonar system for general interaction purposes, we developed finger identification systems by constructing 2D Convolutional deep learning models using sonar fingerprints captured by 16 participants. We achieved the touching finger with peak accuracies of up to 93.7% for both individual and general models and 85.4% for a leave-one-out model.

Second, for user authentication, we developed a multimodal classifier that proposes augmenting smartwatch behavioral biometrics derived from motion, touch, and around-device motion data captured by 24 participants. The results indicate a single thumb tap shows the best performance to authenticate a user with a mean performance of 6.41% average EER, followed by 11.15% for middle finger tap and 13.78% for index finger tap.

1.3 Outlines and contributions

This thesis proposes key contributions with a sonar implementation for an off-the-shelf smartwatch, addressing both general and user-centric interaction challenges. Beginning with a comprehensive overview of previous research, we delve into interaction and user authentication associated with commercial smartwatches. Notably, in **Chapter II**, we offer a detailed examination of smartwatch interactions, emphasizing finger identification and user authentication as drawn from various established studies. Highlighting the limitations posed by the small form factor of smartwatches, we introduce a novel sensing modality for this purpose, Sonar, which is suitable for detecting movement around smart devices without requiring additional hardware.

Chapter III proposes our sonar system for unmodified smartwatches. Our system emits an ultrasonic signal through a speaker located on one side of the smartwatch, and a microphone located on the other side receives it. By processing them using demodulation and auto-correlation, we created sonar fingerprints featuring spatial-temporal information during a touch.

Chapter IV introduces a finger identification system using captured sonar fingerprints during the touch input task on the smartwatch. A description of deep learning models that are capable of using these images to reliably and accurately recognize touching fingers. These data and results indicate that sonar is a promising and effective technique for developing finger identification input systems for smartwatches that do not require modification of existing devices nor rely on users physically instrumenting their touching fingers.

Chapter V implemented and evaluated the first around-device behavioral biometric authentication system for a smartwatch. In addition, this system could be implemented on other platforms equipped with an IMU sensor, a speaker, and a microphone to capture behavioral biometric features.

Chapter VI discuss the feasibility of sonar sensing on unmodified smartwatches based on results of two studies introduced in Chapter IV and V. Furthermore, this section addresses the limitation of both studies and suggests future plans to improve the sonar sensing system for unmodified smartwatches.

In conclusion, through this holistic investigation, the thesis underscores the transformative capability of sonar sensing in unmodified smartwatches that paves the way for more intuitive and secure wearable interactions in the real world.

II Related Work

Smartwatches are reliable, capable, and powerful wearable computers, however, their miniaturized design of smartwatches introduces a unique set of challenges, particularly concerning expressivity and security. The small form factor directly impacts the way users interact with the device, creating limitations that can significantly influence both the range and depth of functionalities smartwatches can offer. To address this issue, we adopted sonar to enable natural and unencumbered sensing on unmodified smartwatches. This section starts by giving an overview of prior works for two different scenarios. The first section summarizes the finger identification system for increasing the expressivity of user interaction from a general user perspective. Then, the second section reviews the behavioral biometric user authentication system from the user-specific angle. In the last section, I discuss why sonar is a promising modality for both finger identification and user authentication on unmodified smartwatches.

2.1 Finger identification input modality

Finger identification has frequently been proposed [33] or evaluated [34] as a mechanism for increasing the expressivity of user interaction on platforms as diverse as tabletop computers [35], smartphones [36] and smartwatches [17, 18]. The core design concept involves assigning different functions to different fingers [37, 38], or sets of fingers [34], a technique that has been demonstrated to be useful in domains as varied as text-entry (by assigning different letters to different fingers) [19] through multi-tasking (by routing input from different fingers to different applications) [39] to command specification in text editing (such as assigning different fingers to copy and paste) [40]. We argue it holds particular value for small devices such as smartwatches.

Finger identification can be implemented via numerous technologies. One approach is to enhance the touch screen, such as by making it capable of detecting fingerprints [41]. Another is to use external sensors, such as a depth camera [42], or a standard camera plus finger-worn visual markers [40, 43]. Other approaches instrument the fingers more invasively, for example, by mounting infrared [44, 45] or vibration [46] sensors on each digit. While these approaches can perform well, they are unsuitable for a real world smartwatch scenario: the small size of watch format devices precludes integration of advanced (and large) touch screen functionality, and wearing additional sensors on the body (e.g., cameras) or fingers is both impractical and undesirable. More practical approaches either use passive finger instrumentation such as a magnetic ring to track fingers [17] or derive data from the detailed analysis of the touch shapes generated by different fingers [18]. While these approaches show promise, both have limitations. Finger flexion can disrupt ring-based magnetic tracking systems [47], and effective performance using touchscreen data is reported to require the adoption of specific and somewhat artificial poses. In order to enable the full potential of the finger identification input modality on wearables, further research on the sensing techniques that can effectively enable it is currently required.

2.2 Behavioral biometrics for user authentication

Smartwatches, increasingly central to accessing sensitive personal data and services, face challenges in user authentication. Traditional methods often struggled on these devices, partly due to the "fat-finger" problem [7, 8] where the user's finger inadvertently obscures or mis-touches the screen. This issue can make authentication cumbersome and error-prone [26]. As a result, users might resort to simpler, easily remembered credentials for convenience [48, 49], compromising security. Unfortunately, such commonly selected passcodes offer only limited security benefits: they are notoriously easy to guess [50]. In addition, if authentication occurs in public spaces, then knowledge-based approaches, such as PIN and pattern, are readily observable by both attackers who may be present in person [51] and by those how have digitally surveilled an area [52].

Biometrics offers potential solutions to these problems: a biometrically secured device can only be unlocked if it recognizes certain characteristics or features of its owner. Such techniques, in the form of fingerprint readers [53] and facial recognition systems [54, 55], have become popular, even dominant [49] solutions for smartphone authentication. Despite their success, these established technologies have rarely been integrated into smartwatches due to the physical (e.g., device surface area occupied), computational or battery resources required. It remains difficult to integrate and power dedicated biometric sensors in a watch form factor device. Consequently, there is growing research interest in using behavioral biometric authentication techniques for smartwatches. These techniques typically re-purpose existing device sensors, such as the touch screen [56] or motion sensors [57], to capture detailed behavioral profiles of users during either sustained or specific interactions. These profiles capture aspects of how users touch, hold, and move their devices that are sufficiently stable and distinct to support effective and accurate authentication. Evidence to support these assertions is mixed: while authentication accuracies can be high (e.g., EER: 1.37% [58]), they are not universally so (e.g., EER: 12.7% [59]). In addition, it may be relatively easy to imitate observed behaviors [60], suggesting these techniques may provide only limited resistance to observation.

2.3 Sonar sensing

Sonar is one promising modality for this purpose. Sonar can be implemented using the speakers and microphones built into smart devices and does not require instrumentation of the touching fingers. On mobile phones, it is well established as an input modality capable of supporting functionality as diverse as the back of device finger tracking [30], mid-air gestures [61], freehand writing [62] and breathing monitoring [31]. To enable this functionality, authors have typically relied on signals from multiple microphones and sought to isolate and track hand or body motion at specific distances and frequencies of interest using approaches such as Orthogonal Frequency Division Multiplexing (OFDM) [28], phase changes in response to multiple continuous waves [61] and impulse response estimation via auto-correlation [30]. A multiple microphone prototype in a smartwatch form factor has also shown the potential of the technique for wearables [28]. However, we know of no sonar implementations for a commercial smartwatch, none that target finger identification or user authentication, and few that rely on data from a single microphone [63]—the hardware configuration available in current watches. The work in this paper seeks to address these omissions and provide insight into the potential of sonar to support finger identification and user authentication using a single microphone on an off-the-shelf smartwatch.

III System

3.1 Sonar system for unmodified smartwatches

Recent commercial smartwatches, such as Galaxy watch 5 [1] and Apple watch series 9 [2], feature a speaker on the left side, a microphone on the right side, and are capable of simultaneously playing and recording 48kHz 16-bit PCM audio. To develop our system, we selected a previously proposed sonar signal: a Zadoff-Chu (ZC) sequence (length 127, $u = 63$, up-sampled to 1024 data points) modulated into a 6kHz band over a 20.25 kHz carrier wave [30]. This signal has a number of beneficial properties. It is, by and large, inaudible. It is brief—just 21.3ms in length—and thus relatively responsive to rapid change (46.95 cycles per second). In addition, it has good auto-correlation properties (i.e., a narrow main lobe and highly attenuated side-lobes), meaning a simple auto-correlation can be used to estimate the impulse response of the signals and distinguish between peaks generated by multiple temporally proximate sonar reflections. It is reported to achieve a spatial accuracy of as low as 3.59mm in the task of tracking index finger location on the back of a smartphone [30]. This demonstrates its suitability for use in interactive systems involving close proximity between a user's hands and sonar sources and receivers. Although prior research has demonstrated the beneficial properties of this signal, we know of no work that has examined its use on a smartwatch nor in a pose recognition task such as finger identification and user authentication.

To achieve pose recognition, our system emits a looping ZC sequence. After a touch occurs, it segments a window of audio around that touch, demodulates the ZC signal, and cyclically estimates its impulse response by performing an auto-correlation against the original 1024 sample ZC sequence data on windows that are 512 samples (10.65ms) apart; this simple manipulation effectively up-samples the sonar measurements we take from 46.95 per second to 93.9 per second [30]. We then generate an image from multiple auto-correlation windows that depicts the changing sonar fingerprint (or impulse response pattern) during a touch. We calibrate the image to focus on the near distance by tracking the largest peak in the impulse response estimations, which is inevitably due to a combination of through-device and direct in-air audio transmis-

[1] https://www.samsung.com/us/watches/galaxy-watch5

[2] https://www.apple.com/apple-watch-series-9

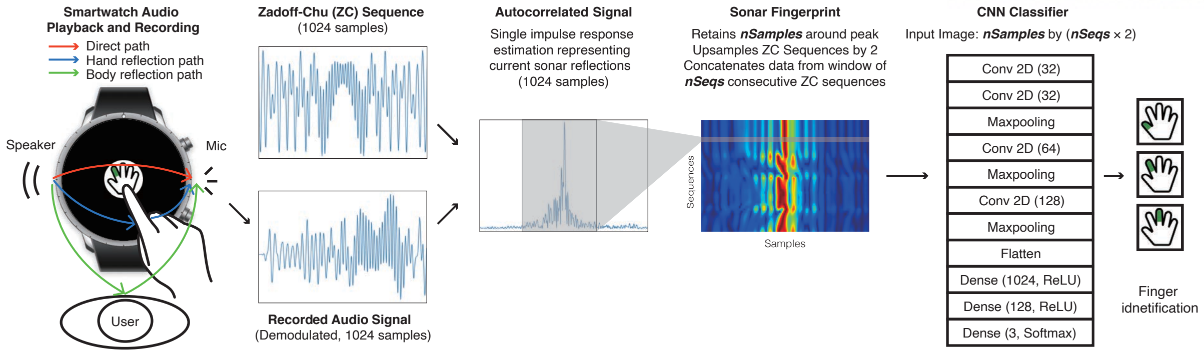

Figure 1: Overview of SonarID: during a screen touch by the thumb, index, or middle finger, a speaker on one side of a smartwatch emits an ultrasonic sonar signal (a Zadoff-Chu (ZC) sequence, modulated over a carrier wave) and a microphone on the other side receives it. The signal is demodulated and processed to create a *sonar fingerprint*: a time-varying image, composed of *nSeqs* ZC sequences, each trimmed to *nSamples* in length, of the impulse response to the signal during the touch. A deep learning model processes this data to identify which finger performed the touch.

sion. Based on this calibration, we trim the sequence to limit our analysis to sonar reflections from predefined ranges. We refer to the two variables in this process as *nSeqs*, or the number of impulse estimations we make, and *nSamples*, or the number of samples from each impulse estimation that we retain. This former variable relates to the amount of data we analyze (e.g., the size of the temporal window used), while the latter measure corresponds to the maximum distance of the sonar reflections we consider.

3.2 Finger identification system

For finger identification system, We implemented SonarID on an unmodified Samsung Galaxy Watch Active 2 smartwatch running Tizen. Figure 1 illustrates the main steps involved in the finger identification processing pipeline. Additionally, all code, scripts, and models in our system are open sources and available for download[3].

Furthermore, figure 2 shows examples of data recorded using this system for no input and for a sequence of three periodic taps with the thumb, index finger, and middle finger made by a single user. Visual examination of these images suggests that the sonar fingerprints generated by these different events are sufficiently unique to support accurate finger classification. All

[3]https://github.com/kjwan4435/SonarID

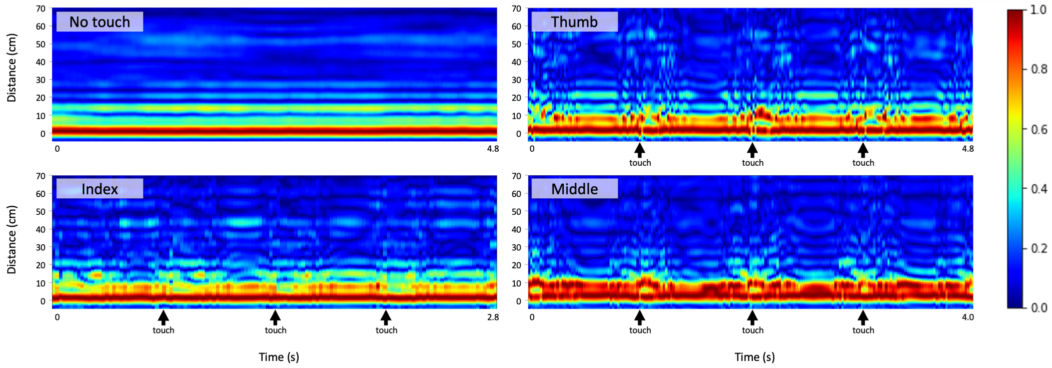

Figure 2: Examples of the sonar fingerprints, or impulse response estimations, generated during periodic smartwatch taps by each finger. Top-left shows no tapping, top-right thumb taps, bottom-left index finger taps, and bottom-right middle finger taps. Finger touches are marked under the axes on each chart. In order to facilitate visual inspection, chart x-axes have different scales (to present taps at the same spatial interval) and y-axes show only half of the auto-correlated signal.

images show a strong immediate response, representing the direct audio path. Index finger taps show limited additional reflections, while the thumb and middle finger taps both present stronger proximate (5 to 10 cm) signal peaks, and distinctive sets of more distant reflections. In addition, the relatively complex, time-varying nature of patterns suggests that approaches to processing the images based on identifying, segmenting, and tracking key signals (e.g., the location of a single object or finger [30]) are poorly suited to the finger identification task. We suggest this is because the key differences in the signals recorded by our system represent the varying surfaces of the hand as it moves through the process of tapping the watch with different fingers. There is no single point of interest; rather, the information is contained within the changing pattern of reflections as a whole. Two key factors likely contribute to these changes: the gross movements of the hand over the watch as it brings the appropriate finger to bear, and the variations in hand pose that touches with different fingers naturally entails—retracting, extending, or holding the various fingers out. Accordingly, rather than try to identify key aspects of the signal, we argue that a more profitable approach will follow recent work in room-level sonar image reconstruction [63] and apply a deep learning approach to the raw sonar data—in our case, the impulse response images.

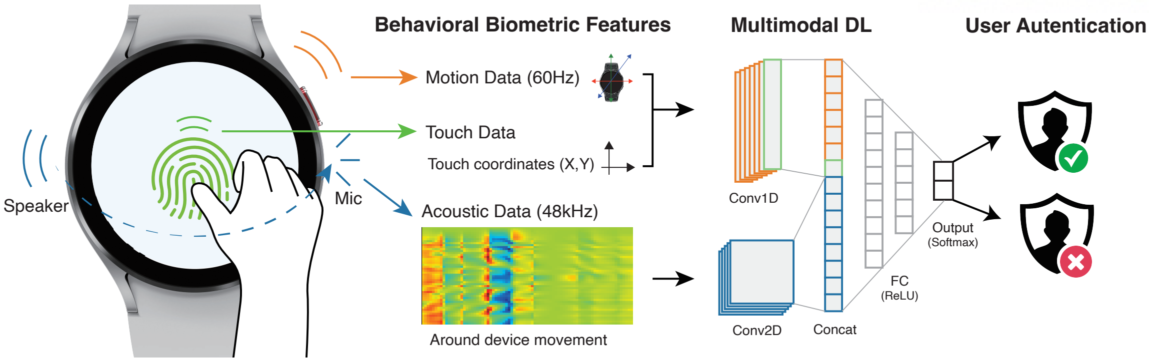

Figure 3: SonarAuth system. A user taps their smartwatch and behavioral biometric features are extracted. In addition to standard touch and motion features, we capture the motions of the hand over the watch during input using active sonar. A structured inaudible ultrasonic signal is emitted from the speaker on one side of the watch, transformed by its passage through the environment, and captured by a microphone on the other side. We process this signal to identify distinctive hand poses and movements that can enhance the performance of smartwatch behavioral biometrics.

3.3 User authentication system

For the user authentication system, we developed SonarAuth as a wearOS application that collects touch coordinates, device movements (accelerometer, gyroscope, magnetometer, gravity, linear acceleration, and rotation vector), and in-air hand motions on an unmodified Samsung Galaxy Watch 4 smartwatch. We capture motion and touch data from the device's screen and IMU at 60 Hz and use its speaker (mounted on the left) and microphone (on the right) to simultaneously play and record 48kHz 16-bit PCM audio sonar signals that represent objects and movements around the device. The sonar system is based on a previously deployed implementation [64]. It uses a looping up-sampled Zadoff-Chu (ZC) sequence (original length $= 127$, $u = 63$, up-sampled to 1024 data points) modulated into a 6kHz band over a 20.25kHz carrier wave. The signal is inaudible and provides excellent auto-correlation features to support the accurate detection of multiple proximate sonar reflections. In addition, each individual ZC sequence spans just 21.3ms, thus ensuring the system remains sensitive to rapid around-device movement. To use SonarAuth, the user simply taps the smartwatch with a finger, and we segment the sensor data in a temporal window around the touch. We then use multimodal deep learning to process this data and authenticate the user. Figure 3 illustrates the overall SonarAuth system.

IV Study 1: Using Sonar to Identify Fingers on a Smartwatch

4.1 Study

We conducted a study to collect sonar data from our system to support the development and evaluation of a deep-learning classifier for finger identification on a smartwatch. In addition, we collected user performance data to inform a characterization of user behaviors that can shed light on the cues a classifier could use to distinguish between touches by different fingers. The study was approved by the local IRB and fully complied with both national laws and institutional regulations related to social distancing.

4.1.1 Participants

Sixteen participants (9 male, all right-handed, mean age of 23.94 (SD 2.32)) were recruited from the local university via online community channels. They were highly familiar with computers (4.93/5, SD 0.25) and smartphones (5/5) but relatively unfamiliar with smartwatches (1.81/5, SD: 1.05). Their hands measured 20.23cm (SD 1.42cm) in length (from base of hand to tip of middle finger), and the mean length of thumb, index, and middle fingers were, respectively, 5.81cm (SD 0.47), 7.09cm (SD 0.49) and 7.74cm (SD 0.48cm). The study took approximately an hour to complete, and each participant was compensated with approximately 13 USD in local currency.

4.1.2 Design

The study was designed around one key variable: fingers. We considered only three fingers, in line with prior work in this area [17]. We specifically selected the thumb, index and middle, as these digits feature in closely related research studying unencumbered finger identification on smartwatches [18] and are commonly used while interacting with pointing devices such as a mobile phone touch screen (thumb and index) or two button mouse (index and middle). In order to capture a range of touches, we laid out seven targets in a circular arrangement: one in

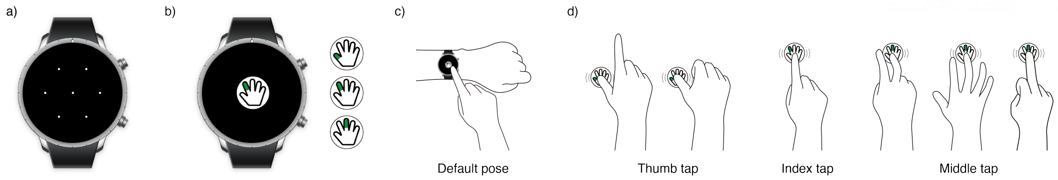

Figure 4: Study interface and interaction. It shows target positions (a), the interface during a trial (b), the index finger pose used for taps to start a trial and during the fixation period (c), and representative examples of different participants' hand posture during thumb, index and middle finger taps (d)

the center, surrounded by the six others. Each target was 120 pixels (11.33 mm) in diameter, one-third of the smartwatch's screen diameter of 34mm. We indicated which finger to use for each screen touch with icons displayed directly on the target. Details of this interface are shown in Figure 4. In total, this arrangement led to 21 different trials, each with a unique finger/target combination. We arranged these trials in four blocks, each featuring a randomly ordered set composed of five repetitions of each possible trial. As such, the study captured data from 6720 trials in total: seven targets by three fingers by five repetitions by four blocks by 16 participants.

For each trial in the study, we captured the following measures: movement-time, measured from the end of the fixation period until touch down over the target; touch-time, the duration the finger was in contact with the screen; correctness, whether or not the appropriate target was selected and; raw audio. This was both emitted and captured from the start of the fixation period until 500ms after the screen was released. In addition, we recorded videos of participants' hands and watch throughout the study and asked them to self-report any erroneously completed (e.g., wrong finger) trials they noted. We did not incorporate any further independent measures of the correctness of tapping fingers. Following prior work in this area [18], we relied on the simplicity of the study task to ensure that the vast majority of the study trials were completed correctly.

4.1.3 Procedure

The experiment took place in an empty classroom with participants seated in front of an empty desk. To prevent fatigue, they rested their wrist on the desk. Participants first read the instructions and signed consent. They then donned the smartwatch on their left wrist and had a maximum of five minutes to practice the three finger touches (thumb, index, middle). They

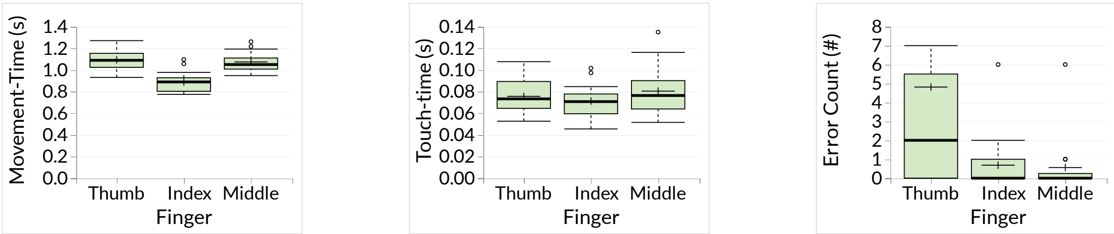

Figure 5: Study data showing movement-times (left), touch-times (center), and error counts (right) for touches with the thumb, index, and middle fingers.

were instructed to determine the most comfortable and effective input actions for completing the study tasks—this was important as many participants were unfamiliar with both smartwatches and the use of either thumb or middle finger to perform taps. This practice stage helped reduce variability in the study tasks as it allowed participants to experiment with different approaches to making the finger taps. However, we note that we did not restrict participants' hand, arm, or finger poses—they were free to tap the watch any way they felt was comfortable. After indicating they were ready, the study began. Each trial involved tapping the screen with the index finger, then hovering over the watch for a fixation period (500ms), after which time the trial target and finger were shown. The participant then tapped the target with the appropriate finger, and the next trial began. Incomplete trials timed out after three seconds. Participants were able to rest at any time between trials, and there were three enforced breaks of two minutes, one between each of the four trial blocks. At the end of the study they completed demographics.

4.2 Behavioral observations

We recorded a total of 6530 (97.17%) correctly completed trials and 97 errors (1.44%) in which the on-screen target was not correctly tapped. In addition, we logged 30 errors (0.45%) in which participants self-reported erroneously tapping the screen (e.g., with the wrong finger) and 63 errors (0.94%) in which technical problems led to data loss. We opted not to analyze self-reported errors due to their sparsity and technical errors due to their lack of relevance to user performance.

Data for movement-time and touch-time from correctly completed trials, and trial error count, are shown in Figure 5. Movement- and touch-time data were normally distributed and upheld sphericity assumptions, so we analyzed them with one-way repeated measures ANOVA on the finger variable. Movement-time led to a significant main effect (F (2, 15) = 123.578, p<0.001, $\hat{\eta}^2_G$=0.493), while touch-time did not (F (2, 15) = 2.715, p=0.082, $\hat{\eta}^2_G$=0.043). Post-hoc t-tests on movement-time, incorporating Bonferroni corrections, indicated that taps with the index finger were faster than with the thumb or middle finger (both p<0.001). This suggests that thumb and middle finger taps in our task both involved additional hand movements prior to contact with the screen and that these motions exacted a modest, but stable, time cost of approximately 200ms. These motions are likely due to the fact that the index finger was mandated for the taps used to start each trial. As prior work has indicated sonar sensing systems on smart devices are highly sensitive to hand and finger motion [30], we note these variations will also likely constitute a key feature that a classifier can use to distinguish between taps by different fingers, with thumb taps involving a descending rightward motion of the hand, index taps a simple downward motion and middle taps a descending leftward motion.

We then examined errors. Error counts were not normally distributed; in fact, three participants contributed 51 (66%) of the errors, with numerous others achieving perfect or near-perfect performance. Additionally, the vast majority of errors (92%) occurred on the middle row of targets—on the leftmost (23%), center (54%), or rightmost (15%) targets. Reflecting the unevenness of this data, we analyzed error counts using a Friedman test on the finger variable. It revealed significant differences ($\chi^2(2)$ = 5.848, p=0.008), which follow up Wilcoxon tests (applying an α level of 0.0167 to emulate Bonferroni correction) indicated were due to the index (Z=2.5, p=0.007) and middle (Z=4.0, p=0.011) finger leading to fewer wrong target selections than the thumb. We conclude that although error rates in the study were generally low (at 1.44% in total), reflecting the simplicity of the study task, there was a subset of participants who experienced some degree of challenge in the task of accurately thumb tapping buttons located in the center row of the watch screen. To support optimal performance for all users, targets intended for thumb tapping might therefore be better located at the top or bottom of a smartwatch screen. We could not infer potential features that might support finger classification from this data due, in part, to its sparsity.

Finally, we took notes live and examined the study video recordings to informally catalog variations in participants' performance of the three taps in the study. Index finger taps were uniformly performed simply with a single, isolated, outstretched digit. For thumb taps, perfor-

Table 1: Accuracy (in %) for different sonar ranges, expressed in *nSamples*, the length of the ZC sequence used. *nSeqs* is set to 60/40.

Signal Range	**Near**		**Mid**		**Far**
nSamples (#)	25	50	100	200	300
Distance (cm)	4.5	9	18	36	54
Accuracy (%)	86.06	89.43	91.73	93.26	93.72

Table 2: Accuracy (in %) for different data capture periods, expressed in terms of both *nSeqs*, the number of impulse estimations performed, and time (ms). In these results, *nSamples* is set to 300.

Signal Period	**Before Touch**			**After Touch**		**Before/After Touch**		
nSeqs (#)	20	40	60	20	40	20/20	40/40	60/40
Time (ms)	213	426	639	213	426	213/213	426/426	639/426
Accuracy (%)	62.94	86.37	89.59	60.95	82.39	69.98	90.81	93.72

mance was also quite uniform, with 14 participants opting to make a "V pose" with the thumb and index, with other fingers tucked into a loose fist. The remaining two participants (P10, P11) performed thumb taps after first retracting their index finger to join the others in their fists. There was somewhat more diversity in middle finger taps, with 11 participants making a "V pose" with index and middle, two participants extending the middle finger alone (again, P10 and P11), and the remaining three (P2, P5, P12) extending all fingers, loosely splaying their hand, during a middle finger tap. These poses are illustrated in Figure 4. While we expected the diversity of these user-selected poses to present challenges to our goal of developing a sonar based recognizer for finger identification, the similarities among many participants' input may also lead to generation of consistent classes of sonar reflection. For most participants, thumb touches meant their hand was situated to the right of the watch, while index touches involved a single digit above the watch and middle touches led to the fingers or hand being on the left of (or covering) the watch. We suggest the patterns of sonar reflections generated by these different poses may be sufficiently distinct to support reliable finger classification.

4.3 Finger identification performance

4.3.1 Preprocessing and classifier

We processed the recorded sound for each trial according to the procedures set out in **Chapter IV**. In brief, the recorded ZC sequence data was demodulated from the carrier wave, then cyclically auto-correlated with the original 1024 sample ZC sequence with a hop size of 512. Each of these auto-correlations yielded an estimate of the impulse response to the signal that encompasses reflections from a maximum distance of approximately 368cm [30]. We concatenated a sequence of these estimations to form a sonar fingerprint—a time-varying image representing the sonar reflections recorded throughout a touch. We defined two variables in this process: *nSamples*, referring to the number of samples used from each impulse response estimation, and *nSeqs*, referring to the number of estimations we concatenate. These correspond to, respectively, the physical range at which we capture sonar reflections and the period of time in which we capture them. We explored the impact of varying the value of these parameters in a grid search procedure (see Tables 1 and 2).

We sought to identify the fingers involved in screen touches from our sonar fingerprints using a 2D convolution neural network (ConvNet) [65]. Due to the similarity between our data and traditional sonograms, we based our system on an existing design for detecting visual patterns in audio spectrograms [66]. To support feature learning, we selected a model composed of three convolution blocks, with each block containing a convolution, Rectified Linear unit (ReLU) activation, and max-pooling layer. We consecutively increased the number of filters (from 32 to 64 then 128) and used a 3x3 kernel in the convolution layers, while we used a 2x2 kernel and a stride length of 2 in the max-pooling layers. After we flattened the output from the last convolution block, we included two fully connected hidden layers with, respectively, 1024 and 128 units. A final fully connected layer performed the multi-classification into thumb, index, or middle finger classes. Figure 1 shows this structure.

4.3.2 Classification performance

We first constructed general models using data from all participants to explore the *nSamples* and *nSeqs* parameters via a grid search procedure. In this process, we used 80% of the data for training and reserved the remaining 20% for final testing. Using the training data set, we conducted five-fold cross-validation procedures. We selected values for *nSamples* to be between 25 to 300, encompassing reflections from between 4.5 cm (single reflections from the finger) to 54 cm (multiple reflections from the body, arm, hand and fingers). We considered *nSeqs* values

Table 3: Accuracy (in %) for individual and LOOCV models, including mean (μ) and standard deviation (σ) from all participants.

Participant	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	μ	σ
Individual (%)	92.3	92.5	93.2	90.9	95.3	94	92.5	95.9	93.3	95.6	93.4	94.9	89	96.6	95.0	94.7	93.7	2
LOOCV (%)	88.3	74.5	89.8	78.3	85.5	81.6	88.6	95.2	85.5	89.1	92	86.3	79.6	80.5	83.4	87.7	85.4	5.4

of between 20 and 100, specifying periods with a duration of between 213ms and 1065ms. To shed light on which periods contain the most salient information, we considered intervals both before the temporal midpoint of a touch (213-639ms), after the temporal midpoint of a touch (213-426ms), and the combination of these ranges. We omitted the 639ms period after the touch mid-point due to the high latency it entails and the fact that we terminated audio capture 500ms after finger up. The performance of models constructed during this grid search procedure on our final test data set are shown in Tables 1 and 2. Perhaps unsurprisingly, they indicate that peak accuracy (93.7%) is achieved with the most data: the configuration that includes the greatest range (*nSamples* set to 300, 54cm reflections) and time period (*nSeqs* set to 100, including 60 samples before and 40 samples after touch midpoint, 1065ms). However, we note that performance with lower distances remains high—data from a range of just 9cm, encompassing the finger and hand, achieves an accuracy of 89.43%. We suggest the benefits from including further ranges may be due, in part, the presence of echoes or multiple reflections from the hand and wrist. In terms of time, periods before the touch midpoint showed modest improvements over those after it, while the combination of both periods led to peaks; we conclude that motions of both finger approach and retraction from the screen contained valuable information to support classification. In addition, the short periods around the touch midpoint, where finger and hand motions are likely small or slow, showed low performance. The key sonar features therefore likely relate to the changing sonar reflections that are recorded as the hand moves. Data from more static poses were less salient. Based on these results, we selected the optimally performing general model configuration, using data from a range of 54cm (*nSamples* $=$ 300) captured over periods of 1065ms (*nSeqs* $=$ 100), for all further tests.

To validate performance, we produced two further sets of models: individual models and Leave One Out Cross Validation (LOOCV) models. In the individual models, each participant's own data was used to produce a model. We followed the same procedures used to create the general model: an 8:2 train/test data split and five-fold cross-validation procedures on the

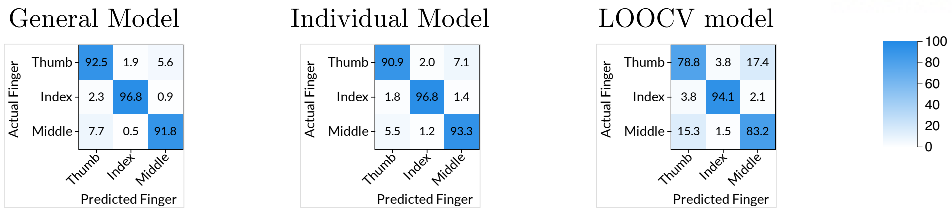

Figure 6: Confusion matrices for SonarID classifiers (% accuracy). Left: general model; center: mean individual model; right: mean LOOCV model.

training set. For the LOOCV models, data from each participant served as a final test set for a model trained using five-fold cross-validation on data from all other participants (essentially, a 15:1 train/test split). Data from these models are shown in Table 3. Individual models showed a similar performance profile to the general model: mean accuracy was also 93.7%. We suggest that individual models were unable to improve on the general model performance due to the relatively sparsity of data in each. Data from the LOOCV models reinforces this conclusion. While fair (85.37%), it is notably reduced from that achieved using the general or per-user models. This suggests that, within our relatively small sample, the sonar fingerprints created by each users' tapping behavior were somewhat specific to that user. In order to achieve LOOCV performance equivalent to that in the general model, or individual model performance that exceeds it, we would likely need to sample more data from more users.

It is also worth reflecting on the distribution of classification errors among the finger classes. Figure 6 shows this for the general model and the mean performance for the sets of both individual and LOOCV models. Similar trends can be observed. Classification performance peaks with the index finger while the thumb and middle are more frequently confused. This effect is particularly prominent in the LOOCV models, where the thumb and middle fingers are correctly recognized just 78.8% and 83.2% of the time. We can extract a number of implications from this result. First and foremost, it suggests that performance in a two-class task [19] of separating the index taps from those of other fingers would be high: up to 96% even in the LOOCV models. Secondly, it likely reflects the diversity of strategies we observed for middle and thumb taps—while index finger taps were universal, there were several different high-level strategies (such as tucking in or extending unused fingers) during thumb and middle taps. This diversity may have increased the difficulty of the classification task. As such, larger training sets may be able to more clearly separate finger touches based on various different strategies.

Alternatively, customized user models may be able to more accurately cater to a given user's particular touching style.

V Study 2: Using Sonar to Improve Behavioral Biometrics on a Smartwatch

5.1 Study

We conducted a study to collect behavioral biometric data to explore the feasibility and performance of SonarAuth. In addition to logging sensor watch data, we recorded videos of the tapping behaviors of each participant. The study was approved by the local IRB.

5.1.1 Participants

We recruited 24 participants (12 males, mean age of 22.13 (SD: 3.57)) from the local university. They are all right-handed, and relatively unfamiliar with smartwatches (2.13/5, SD: 1.42), compared to smartphones (5/5) or computers(4.67/5, SD:0.56). Participants' hand span was a mean of 18.5cm (SD:2.88cm), their hand length was a mean of 18.2cm (SD:1.3cm), and their palm width was a mean of 7.75cm (SD:0.9cm).

5.1.2 Experiment protocol

The experiment took place in an empty classroom with participants seated on a fixed chair in front of an empty desk. Before the experiment, participants read instructions and signed informed consent. After wearing the smartwatch on their left (non-dominant) wrist, participants completed a short practice session, in which they were asked to make a total of 30 taps to the watch, practicing this action with each of their thumb, index, and middle fingers. They were instructed that the goal of this session was to arrive at a comfortable and efficient way of performing each finger tap. After completing this session, the formal study began.

The study was composed of eight blocks of 30 trials, with ten trials in each block requiring the use of one of the thumb, index, or middle fingers. Trials in each block were randomly ordered. Each trial commenced by requesting participants tap on an on-screen button (with any finger) to

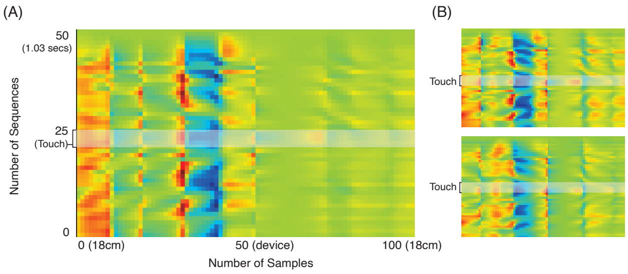

Figure 7: Examples of generated sonar image, showing the impulse response estimate from a smartwatch touch (A), and two examples of augmented data derived from the same signal using the phase rotation technique (B).

begin. A centrally located trial target was then presented, and it visually depicted which finger to use. The participant's task was simply to tap it with the appropriate digit. The trial then ended and the next trial began. We asked participants to verbally report any errors (e.g., wrong finger use) in order to facilitate excluding such data from subsequent analysis. Participants were able to rest at any time between trials. Between blocks, they were required to take off and re-wear the smartwatch—this helped ensure the data captured reflected the kind of minor variations in wearing pose that would occur naturally during daily device use. At the end of the study, participants completed demographics and the NASA Task Load Index (TLX), a measure designed to assess the workload experienced during the use of SonarAuth.

5.2 User authentication performance

In total, data from 5760 trials were recorded: 10 repetitions by three fingers by eight blocks by 24 participants. Participants self-reported 11 errors (0.19%) in which they used the wrong finger to tap the targets and these trials were excluded from further analysis. For each of the remaining 5749 trials, we captured and analyzed touch coordinates and motion data at 60Hz and the audio signal containing the reflected structured inaudible ultrasonic signal.

5.2.1 Preprocessing and multimodal classifier

To analyze the captured audio signals, we demodulate the signal and estimate its impulse response by conducting auto-correlation against the original up-sampled ZC sequences. By concatenating the results of auto-correlation, we can generate an image illustrating the changing

Table 4: Mean of individual user authentication performances for each finger tap in both multimodal and sonar image classifiers.

Classifier	Finger	EER (%) Mean (SD)	FAR (%) Mean (SD)	FRR (%) Mean (SD)
Multimodal	Thumb	6.41 (5.1)	8.02 (8.4)	4.71(5.3)
	Index	13.78(7.6)	13.94(13.1)	16.55(12.3)
	Middle	11.15(6.4)	9.57(6.4)	14.27(11.6)
Sonar Image	Thumb	14.6(9.03)	12.4(8)	16.1(11.2)
	Index	19.26(11.2)	17.21(9.3)	21.21(15.8)
	Middle	15.77(8.8)	13.6(7.7)	16.82(11.9)

sonar response during the authentication tap. To avoid including extraneous data, we generate the image using ZC sequences captured immediately around the screen touch. Specifically, we capture 50 ZC sequences of data (25 before and after the finger touches the surface), representing a 1.065-second period (50 by 21.3ms) around the finger-down event. In addition, we limit the distances at which we track sonar signals to spatial regions immediately around the watch by considering the auto-correlation data only around its highest peak (representing the direct signal path between the speaker and microphone). By considering 100 samples (50 before and after the peak points), we capture data from approximately 18cm around the watch (343(m/s) * 0.021(ms) * 50(samples) / 2(reflections)). We then remove static components of the image by subtracting the first signal from all subsequent signals. Finally, we conduct data augmentation to increase our data set size. Specifically, we rely on the fact that the asynchronous speaker and microphone generate an unpredictable phase offset [67] and generated eight different sonar images from data for each tap by rotating the phase of the ZC sequences.

To analyze the motion data and touch coordinates, we first merged the data from different sensors based on the closest timestamps. In the periods before and after touch, we padded the touch coordinate data with zeros. We also windowed the data for one second around the touch-down by retaining 60 samples (30 before and after the touch-down). Finally, all motion data was normalized.

To sum up, we generated 8 augmented sonar images (50: the number of sequences, 100: the number of samples) with motion and touch biometric features (60: the number of data points, 20: the number of features) for each tapping trial.

We developed a multimodal classifier to combine motion, touch, and sonar images. For motion and touch data, we built a simple 1D CNN classifier consisting of two 1D convolutional layers (the number of filters: 128, 256 each, kernel size: 10, stride: 2, ReLU activation) and each is followed by max-pooling layers (pool size: 2) and a dropout layer (p: 0.5). For sonar images, we built a 2D CNN classifier consisting of four 2D convolutional layers (the number of filters: 32, 64, 128, 256 each, kernel size: (3,3), stride: (1,1), ReLU activation) with max-pooling layers(pool size: (2,2), stride: (2,2)). After flattening the output from the last convolution block from both uni-modal classifiers, we concatenate them and feed them into three fully connected dense layers with, respectively, 1024, 512, and 128 units. To validate our multimodal classifier, we also evaluate the efficacy of a single 2D CNN classifier for sonar images (directly connected to a series of fully connected dense layers after flattening). As we evaluate our system on three fingers for every user, we constructed a total of 72 (3 fingers * 24 users) authentication models. In every model, except for a genuine user, all other 23 participants are labeled as attackers. We then randomly split 23 attackers into 17 for training and 6 for testing the model. Among 8 blocks of each user's data, the first 6 blocks are used for training and the last 2 blocks are used for testing. We balanced the number of data between genuine users and attackers by randomly selecting a subset of data from each attacker during training.

5.2.2 Classification performance

To evaluate our classifiers, we measure Equal Error Rate (EER), False Acceptance Rate (FAR), and False Rejection Rate (FRR). Table 4 shows overall performance in both multimodal and sonar image classifiers. The multimodal classifier shows stronger performance for the thumb tap (Mean EER: 6.41 (SD 5.1), FAR: 8.02, FRR: 4.71), followed by the middle finger tap (Mean EER: 11.15 (SD 6.4), FAR: 9.57, FRR: 14.27) and worst performance with the index tap (Mean EER: 13.78 (SD 7.6), FAR: 13.94, FRR: 16.55). The sonar image classifier shows a similar trend. Peak performance is achieved with the thumb tap (Mean EER: 14.6 (SD 9.03), FAR: 12.4, FRR: 16.1) then the middle finger tap (Mean EER: 15.77 (SD 8.8), FAR: 13.6, FRR: 16.82) and finally the index tap (Mean EER: 19.26 (SD 11.2), FAR: 17.21, FRR: 21.21). Overall, the multimodal classifier shows marked improvements in performance over those achieved with the sonar image classifier.

Table 5: Usability data in terms of ratings of performance, comfort, and speed across thumb, index, and middle finger taps

	preference(%)			comfort(%)			speed(%)		
	Thumb	Index	Middle	Thumb	Index	Middle	Thumb	Index	Middle
1st	12.5	87.5	0	16.7	79.17	4.2	12.5	79.17	8.3
2nd	20.8	4.17	75	20.8	16.67	62.5	20.8	16.67	62.5
3rd	66.7	8.33	25	62.5	4.17	33.3	66.7	4.17	29.2

5.3 Usability

After the study, participants reported usability in terms of preference, comfort, and speed regarding the three finger-tap; thumb, index, and middle tap. By and large, participants marked index tap as the first in all metrics (preference: 87.5%, comfort: 79.17%, speed: 79.17%) followed by thumb tap (preference: 12.5%, comfort: 16.7%, speed: 12.5%) and middle tap (preference: 0%, comfort:4.2%, speed: 8.3%). Although the thumb had its adherents, we also note that it was also most frequently rated as the least preferred option (preference: 66.7%, comfort: 62.5%, speed: 66.7%), suggesting participants were split in their subjective opinions of this input action. Table 5 shows the full results from this survey. Participants also completed the NASA-TLX to report their mental workload. The results were as follows: mental demand (5.17/20, SD:5.04), physical demand (4.21/20, SD:4.24), temporal demand (9.67/20, SD:6.11), performance achieved (4.5/20, SD:6.14), effort expended (5.96/20, SD:5.3), and frustration experienced (2.58/20, SD:3.63).

VI Discussion

This thesis explored the use of sonar sensing on unmodified smartwatches. Our focus centered around not just understanding the feasibility of sonar sensing but also unraveling its potential applications in user interaction and authentication.

At the core of this exploration was that unmodified smartwatches, though limited in screen space, could potentially harness their built-in hardware for novel interaction paradigms. In **Chapter IV**, by using sonar sensing, SonarID aimed to identify finger-based interactions, offering a new dimension of input modality beyond the traditional touch. The results and implications of this study are promising. Beyond just broadening interaction techniques, such methodologies have the potential to offer personalized experiences, recognizing users by their unique interaction signatures. Building on the foundational principles of SonarID, in **Chapter V**, SonarAuth sought to push the boundaries further by utilizing sonar sensing as a mechanism for behavioral biometrics. The premise was clear – if sonar could be used to identify fingers, could it also be employed to authenticate users based on unique tapping behaviors? The results were intriguing. Though there is room for improvement, the concept of single-tap authentication harnessing the power of sonar paints an optimistic future for wearable device security.

6.1 Feasibility of sonar sensing on smartwatches

Chapter IV presents SonarID, a system that recognizes the finger tapping an off-the-shelf smartwatch using a sonar scene generated and sensed via the device's built-in speaker and microphone. It achieves a peak accuracy of 93.7% for both individual and general models and a LOOCV accuracy of 85.4%. These results compare favourably to prior work implementing finger identification using standard smartwatches. TriTap [18], for example, uses touchscreen data for a similar classification task and achieves an accuracy of 79.4% using individual models and natural touches. Furthermore, the performance we report is only marginally reduced compared to systems that track worn objects, such as the magnetic ring in MagTouch [17] that supports a recognition rate (among index, middle, and ring fingers) of 95.03%. We argue that the convenience and practicality of enabling input with an unencumbered touching hand offer advantages over systems that require users to wear additional hardware.

It is also worth discussing the detailed performance profile of our results. We show strong performance in classifying the index finger and difficulty distinguishing between thumb and middle. TriTap's [18] capacitive screen implementation shows the strongest performance in distinguishing the thumb and confuses the index and middle. A multi-modal combination of these two approaches would likely be highly complementary. Similarly, SonarID's ability to classify thumb taps may be able to extend MagTouch's [17] finger-only approach with an additional digit. We see strong potential in combining these modalities in the future. Data from the grid search over temporal periods for capturing a sonar signal (*nSeqs*) also presents implications for the design of interactive systems. It suggests that peak performance will require continual emission and recording of sound—the highest performance we observed uses data from both before and after touch. Although prior work has argued that power consumption for sonar systems is reasonable on a smartphone [30], the impact on smartwatch battery life may be more extreme—during the intensive watch use in our hour long study, watch battery level declined by 8%-10% for each participant. A more efficient approach of recording signals only after a touch offers reduced classification accuracy (of 82.39%), suggesting a potential trade-off between power consumption and classification performance. We also note that using sonar data from after a touch implies a latency (roughly equivalent, in our studies, to the 500ms commonly used for long tap) before events can be detected. While that latency can be avoided by relying on data gathered up until a touch occurs, this approach also leads to a more modest reduction in classification accuracy (to 89.59%). Future studies and system designs will need to flesh out and balance these concerns.

Chapter V proposes SonarAuth [68], a user authentication system based on behavioral biometrics derived from motion, touch, and sonar-sensed around-device movement on an unmodified smartwatch. Based on data from a single tap on a smartwatch screen, we achieved a mean performance of 6.41% EER for thumb taps, followed by 11.15% EER for middle finger taps and 13.78% EER for index finger taps (13.78% average EER) using a multimodal classifier. A classifier using only sonar data also shows a similar trend but achieves approximately 6% lower performances overall. This suggests that a multimodal approach improves behavioral biometric features on smartwatches for user authentication.

We can make a number of other interesting observations. Perhaps most importantly, we can shed light on how the thumb tap outperforms the index and middle fingers by examining and analyzing participants' tapping behaviors, as recorded in video of the study. There are several factors that may contribute to the increased performance of the thumb. For instance, as the thumb is shorter than the other fingers, participants' hands must be located closer to the watch during thumb screen taps, which may provide more information related to the around-device movement. In addition, we note that participants showed relatively similar tapping hand poses for the index and middle finger taps (based on extending the tapping digit alone) compared to the thumb tap (which showed various articulations of the other fingers) in recorded videos. This may have made the hand poses used during thumb taps more distinctive. Participants also reported the thumb tap is the least preferable, a result that may be due to the requirement for larger gross movements of the hand. However, more positively, we note that NASA-TLX result shows that single tapping authentication, in general, places very low burdens on users in terms of mental and physical demand [69].

6.2 Limitations and future works

There are a number of further limitations to this work; these signpost the next steps for this project.

In the first study, the simplest relates to the abundancy of the data we rely on—results from both our individual and LOOCV models suggest that improved performance could be achieved with larger training sets. While this could be achieved simply via extended empirical work, we also highlight the need to explore data augmentation techniques on our existing data set [66]. While such techniques may enable us to improve the performance of our classifiers, we also note further studies are inevitable. Although much work in finger identification uses a single study [18] or pose [19], other authors note the advantages in terms of robustness and validity that can be realized by sampling data from various situations [17]. While the seated, hand-ready pose we use in this work is both common and representative, a clear next step for this work is to capture data from more diverse situations and environments. These should include while standing, with the arms in various poses, and with an increased diversity of tapping styles, such as fully separated single touches that each involve a finger approaching the watch independently. While performance may change if data is captured from more diverse poses, we note that our current results indicate that proximate sonar reflections from 4.5cm and 9cm, involving just the touching fingers and hand (see Table 1), lead to relatively high accuracy levels of up to 89.43%. This result suggests SonarID may be resilient to changes in upper arm or body pose as such variations would likely impact only more distant sonar reflections.

In the second study, the peak result, 6.41% average EER for thumb tap, suggests our system is not sufficient to support effective user authentication. However, as a series of taps could be concatenated into a sequence, we note it may be possible to increase performance in future work. In addition, this paper validates the classifiers we train without separate multi-recall sessions where diverse variations in tapping behavior may occur— more prolonged empirical work is required to establish if the performance we report for SonarAuth holds in the wild.

In addition, it will be important to explore the robustness of our technique to various forms of environmental disturbance, such as different ambient noise conditions (situations in which closely related prior sonar systems have performed well [30]), or wind—a form of environmental disturbance widely acknowledged to reduce the signal to noise ratio of sonar systems. Our future plans for this project involve addressing these issues. By capturing more data in diverse settings and exploring techniques to augment that data, we believe we can construct classifiers with improved accuracy and increased validity. Doing so will ensure the technique we describe works not only on off-the-shelf devices but also in real-world settings.

VII Conclusion

This thesis demonstrates the feasibility of sonar sensing in unmodified smartwatches, focusing on the transformative potential of utilizing sonar sensing for two primary objectives: general interaction and user-specific perspectives.

The first study explores the general interaction perspective, a finger identification system. This system proposes a natural and unencumbered finger identification on commercial smartwatches by processing sonar scenes recorded during a touch. By training deep learning recognizer using data captured from 16 participants, I achieved a high accuracy of finger recognition with a peak accuracy rate of 93.7% for both individual and general models and 85.4% for the LOOCV models. This result suggests sonar sensing is a robust and efficient solution for finger identification without requiring any additional hardware. The second study pivots to a user-specific angle, user authentication. In this study, I suggest utilizing sonar sensing to detect around-device movement to improve smartwatch behavioral biometrics for mobile authentication. I trained a multimodal deep-learning classifier using motion, touch, and around-device motion data captured from 24 participants. In this result, Tapping with the thumb showed the most favorable results, achieving an Equal Error Rate (EER) of just 6.41%, a promising level of performance from almost the simplest possible input action—just a tap to the screen.

While both studies have paved the way, the broader argument this thesis presents is the untapped potential of sonar sensing in wearables. As technology continues to evolve with further refinement, sonar sensing could emerge as the gold standard in wearable device security and user interaction, combining ease of use with reliability in real-world settings.

References

[1] T. Mishra, M. Wang, A. A. Metwally, G. K. Bogu, A. W. Brooks, A. Bahmani, A. Alavi, A. Celli, E. Higgs, O. Dagan-Rosenfeld *et al.*, "Pre-symptomatic detection of covid-19 from smartwatch data," *Nature biomedical engineering*, vol. 4, no. 12, pp. 1208–1220, 2020.

[2] T. Hao, C. Bi, G. Xing, R. Chan, and L. Tu, "Mindfulwatch: A smartwatch-based system for real-time respiration monitoring during meditation," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, Sep. 2017. [Online]. Available: https://doi.org/10.1145/3130922

[3] Y. Ha, M. Karyda, and A. Lucero, "Exploring virtual rewards in real life: A gimmick or a motivational tool for promoting physical activity?" in *Proceedings of the 2020 ACM Designing Interactive Systems Conference.* New York, NY, USA: Association for Computing Machinery, 2020, p. 1847–1858. [Online]. Available: https://doi.org/10.1145/3357236.3395477

[4] C. X. Lu, B. Du, H. Wen, S. Wang, A. Markham, I. Martinovic, Y. Shen, and N. Trigoni, "Snoopy: Sniffing your smartwatch passwords via deep sequence learning," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 4, Jan. 2018. [Online]. Available: https://doi.org/10.1145/3161196

[5] A. Visuri, N. van Berkel, J. Goncalves, R. Rawassizadeh, D. Ferreira, and V. Kostakos, "Understanding usage style transformation during long-term smartwatch use," *Personal and Ubiquitous Computing*, vol. 25, pp. 535–549, 2021.

[6] M. Perebner, H. Huang, and G. Gartner, "Applying user-centred design for smartwatch-based pedestrian navigation system," *Journal of location based services*, vol. 13, no. 3, pp. 213–237, 2019.

[7] Y. Ren and A. S. Arif, “Investigating a force-based selection method for smartwatches in a 1d fitts’ law study and two new character-level keyboards,” in *Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI ’23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3569009.3572741

[8] H. V. Le, “Fully hand-and-finger-aware smartphone interaction,” in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–4. [Online]. Available: https://doi.org/10.1145/3170427.3173023

[9] A. Butler, S. Izadi, and S. Hodges, “Sidesight: Multi-"touch" interaction around small devices,” in *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’08. New York, NY, USA: Association for Computing Machinery, 2008, p. 201–204. [Online]. Available: https://doi.org/10.1145/1449715.1449746

[10] J. McIntosh, P. Strohmeier, J. Knibbe, S. Boring, and K. Hornbæk, “Magnetips: Combining fingertip tracking and haptic feedback for around-device interaction,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* New York, NY, USA: Association for Computing Machinery, 2019, p. 1–12. [Online]. Available: https://doi.org/10.1145/3290605.3300638

[11] H. Wen, J. Ramos Rojas, and A. K. Dey, “Serendipity: Finger gesture recognition using an off-the-shelf smartwatch,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* New York, NY, USA: Association for Computing Machinery, 2016, p. 3847–3851. [Online]. Available: https://doi.org/10.1145/2858036.2858466

[12] Y. Zhang, T. Gu, C. Luo, V. Kostakos, and A. Seneviratne, “Findroidhr: Smartwatch gesture input with optical heartrate monitor,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, Mar. 2018. [Online]. Available: https://doi.org/10.1145/3191788

[13] M.-C. Hsiu, D.-Y. Huang, C. A. Chen, Y.-C. Lin, Y.-p. Hung, D.-N. Yang, and M. Chen, “Forceboard: Using force as input technique on size-limited soft keyboard,” in *Proceedings of the 18th International Conference on Human-Computer Interaction*

with Mobile Devices and Services Adjunct, ser. MobileHCI '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 599–604. [Online]. Available: https://doi.org/10.1145/2957265.2961827

[14] I. Oakley, C. Lindahl, K. Le, D. Lee, and M. R. Islam, "The flat finger: Exploring area touches on smartwatches," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2016, p. 4238–4249. [Online]. Available: https://doi.org/10.1145/2858036.2858179

[15] B. Lafreniere, C. Gutwin, A. Cockburn, and T. Grossman, "Faster command selection on touchscreen watches," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2016, p. 4663–4674. [Online]. Available: https://doi.org/10.1145/2858036.2858166

[16] I. Oakley, D. Lee, M. R. Islam, and A. Esteves, "Beats: Tapping gestures for smart watches," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2015, p. 1237–1246. [Online]. Available: https://doi.org/10.1145/2702123.2702226

[17] K. Park, D. Kim, S. Heo, and G. Lee, "Magtouch: Robust finger identification for a smartwatch using a magnet ring and a built-in magnetometer," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–13. [Online]. Available: https://doi.org/10.1145/3313831.3376234

[18] H. Gil, D. Lee, S. Im, and I. Oakley, "Tritap: Identifying finger touches on smartwatches," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2017, p. 3879–3890. [Online]. Available: https://doi.org/10.1145/3025453.3025561

[19] A. Gupta and R. Balakrishnan, "Dualkey: Miniature screen text entry via finger identification," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2016, p. 59–70. [Online]. Available: https://doi.org/10.1145/2858036.2858052

[20] P. Rao Gadahad and A. Joshi, "Wearable activity trackers in managing routine health and fitness of indian older adults: Exploring barriers to usage," in *Nordic Human-Computer Interaction Conference*, ser. NordiCHI '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: https://doi.org/10.1145/3546155.3546645

[21] E. A. Ankrah, F. L. Cibrian, L. M. Silva, A. Tavakoulnia, J. A. Beltran, S. E. Schuck, K. D. Lakes, and G. R. Hayes, "Me, my health, and my watch: How children with adhd understand smartwatch health data," *ACM Trans. Comput.-Hum. Interact.*, dec 2022, just Accepted. [Online]. Available: https://doi.org/10.1145/3577008

[22] S. Panda, Y. Feng, S. G. Kulkarni, K. K. Ramakrishnan, N. Duffield, and L. N. Bhuyan, "Smartwatch: Accurate traffic analysis and flow-state tracking for intrusion prevention using smartnics," in *Proceedings of the 17th International Conference on Emerging Networking EXperiments and Technologies*, ser. CoNEXT '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 60–75. [Online]. Available: https://doi.org/10.1145/3485983.3494861

[23] U. Kishnani, N. Noah, S. Das, and R. Dewri, "Privacy and security evaluation of mobile payment applications through user-generated reviews," in *Proceedings of the 21st Workshop on Privacy in the Electronic Society*, ser. WPES'22. New York, NY, USA: Association for Computing Machinery, 2022, p. 159–173. [Online]. Available: https://doi.org/10.1145/3559613.3563196

[24] C. X. Lu, B. Du, H. Wen, S. Wang, A. Markham, I. Martinovic, Y. Shen, and N. Trigoni, "Snoopy: Sniffing your smartwatch passwords via deep sequence learning," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 4, jan 2018. [Online]. Available: https://doi.org/10.1145/3161196

[25] R. Mayrhofer and S. Sigg, "Adversary models for mobile device authentication," *ACM Comput. Surv.*, vol. 54, no. 9, oct 2021. [Online]. Available: https://doi.org/10.1145/3477601

[26] G. Rakhmetulla and A. S. Arif, "Crownboard: A one-finger crown-based smartwatch keyboard for users with limited dexterity," in *Proceedings of the 2023 CHI Conference on*

Human Factors in Computing Systems, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3544548.3580770

[27] J. Lee, J. Kwon, and H. Kim, "Reducing distraction of smartwatch users with deep learning," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, ser. MobileHCI '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 948–953. [Online]. Available: https://doi.org/10.1145/2957265.2962662

[28] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "Fingerio: Using active sonar for fine-grained finger tracking," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2016, p. 1515–1525. [Online]. Available: https://doi.org/10.1145/2858036.2858580

[29] N. Kim, J. Lee, J. J. Whang, and J. Lee, "Smartgrip: grip sensing system for commodity mobile devices through sound signals," *Personal and Ubiquitous Computing*, vol. 24, no. 5, pp. 643–654, 2020.

[30] K. Sun, T. Zhao, W. Wang, and L. Xie, "Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 591–605. [Online]. Available: https://doi.org/10.1145/3241539.3241568

[31] X. Song, B. Yang, G. Yang, R. Chen, E. Forno, W. Chen, and W. Gao, "Spirosonic: Monitoring human lung function via acoustic sensing on commodity smartphones," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3372224.3419209

[32] R. Nandakumar, S. Gollakota, and J. E. Sunshine, "Opioid overdose detection using smartphones," *Science translational medicine*, vol. 11, no. 474, 2019.

[33] Q. Roy, Y. Guiard, G. Bailly, É. Lecolinet, and O. Rioul, "Glass+skin: An empirical evaluation of the added value of finger identification to basic single-touch interaction on touch screens," in *Human-Computer Interaction – INTERACT 2015*, J. Abascal, S. Barbosa,

M. Fetter, T. Gross, P. Palanque, and M. Winckler, Eds. Cham: Springer International Publishing, 2015, pp. 55–71.

[34] A. Goguey, M. Nancel, G. Casiez, and D. Vogel, "The performance and preference of different fingers and chords for pointing, dragging, and object transformation," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* New York, NY, USA: Association for Computing Machinery, 2016, p. 4250–4261. [Online]. Available: https://doi.org/10.1145/2858036.2858194

[35] C. Harrison, J. Schwarz, and S. E. Hudson, "Tapsense: Enhancing finger interaction on touch surfaces," in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 627–636. [Online]. Available: https://doi.org/10.1145/2047196.2047279

[36] H. V. Le, S. Mayer, and N. Henze, "Investigating the feasibility of finger identification on capacitive touchscreens using deep learning," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, ser. IUI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 637–649. [Online]. Available: https://doi.org/10.1145/3301275.3302295

[37] A. Sugiura and Y. Koseki, "A user interface using fingerprint recognition: Holding commands and data objects on fingers," in *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '98. New York, NY, USA: Association for Computing Machinery, 1998, p. 71–79. [Online]. Available: https://doi.org/10.1145/288392.288575

[38] J. Zheng and D. Vogel, "Finger-aware shortcuts," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* New York, NY, USA: Association for Computing Machinery, 2016, p. 4274–4285. [Online]. Available: https://doi.org/10.1145/2858036.2858355

[39] A. Gupta, M. Anwar, and R. Balakrishnan, "Porous interfaces for small screen multitasking using finger identification," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, ser. UIST '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 145–156. [Online]. Available: https://doi.org/10.1145/2984511.2984557

[40] A. Goguey, G. Casiez, D. Vogel, F. Chevalier, T. Pietrzak, and N. Roussel, “A three-step interaction pattern for improving discoverability in finger identification techniques,” in *Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST’14 Adjunct. New York, NY, USA: Association for Computing Machinery, 2014, p. 33–34. [Online]. Available: https://doi.org/10.1145/2658779.2659100

[41] C. Holz and P. Baudisch, “Fiberio: A touchscreen that senses fingerprints,” in *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’13. New York, NY, USA: Association for Computing Machinery, 2013, p. 41–50. [Online]. Available: https://doi.org/10.1145/2501988.2502021

[42] S. Sridhar, A. Markussen, A. Oulasvirta, C. Theobalt, and S. Boring, “Watchsense: On- and above-skin input sensing through a wearable depth sensor,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2017, p. 3891–3902. [Online]. Available: https://doi.org/10.1145/3025453.3026005

[43] J. Wang and J. Canny, “Fingersense: Augmenting expressiveness to physical pushing button by fingertip identification,” in *CHI ’04 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA ’04. New York, NY, USA: Association for Computing Machinery, 2004, p. 1267–1270. [Online]. Available: https://doi.org/10.1145/985921.986040

[44] P. Ewerling, A. Kulik, and B. Froehlich, “Finger and hand detection for multi-touch interfaces based on maximally stable extremal regions,” in *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces*, ser. ITS ’12. New York, NY, USA: Association for Computing Machinery, 2012, p. 173–182. [Online]. Available: https://doi.org/10.1145/2396636.2396663

[45] S. Murugappan, Vinayak, N. Elmqvist, and K. Ramani, “Extended multitouch: Recovering touch posture and differentiating users using a depth camera,” in *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST ’12. New York, NY, USA: Association for Computing Machinery, 2012, p. 487–496. [Online]. Available: https://doi.org/10.1145/2380116.2380177

[46] D. Masson, A. Goguey, S. Malacria, and G. Casiez, "Whichfingers: Identifying fingers on touch surfaces and keyboards using vibration sensors," in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 41–48. [Online]. Available: https://doi.org/10.1145/3126594.3126619

[47] D. Ashbrook, P. Baudisch, and S. White, "Nenya: Subtle and eyes-free mobile input with a magnetically-tracked finger ring," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2011, p. 2043–2046. [Online]. Available: https://doi.org/10.1145/1978942.1979238

[48] F. Sterk and A. Heinemann, "It is not as simple as that: Playing out password security trainings in order to nudge password changes," in *European Interdisciplinary Cybersecurity Conference*, ser. EICC. New York, NY, USA: Association for Computing Machinery, 2021, p. 20–25. [Online]. Available: https://doi.org/10.1145/3487405.3487653

[49] T. Anusas-amornkul, "Strengthening password authentication using keystroke dynamics and smartphone sensors," in *Proceedings of the 9th International Conference on Information Communication and Management*, ser. ICICM 2019. New York, NY, USA: Association for Computing Machinery, 2019, p. 70–74. [Online]. Available: https://doi.org/10.1145/3357419.3357425

[50] Z. Li, T. Li, and F. Zhu, "An online password guessing method based on big data," in *Proceedings of the 2019 3rd International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence*, ser. ISMSI 2019. New York, NY, USA: Association for Computing Machinery, 2019, p. 59–62. [Online]. Available: https://doi.org/10.1145/3325773.3325779

[51] T.-C. Yu, S.-Y. Fang, H.-S. Chiu, K.-S. Hu, P. H.-Y. Tai, C. C.-F. Shen, and H. Sheng, "Pin accessibility prediction and optimization with deep learning-based pin pattern recognition," in *Proceedings of the 56th Annual Design Automation Conference 2019*, ser. DAC '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3316781.3317882

[52] S. Siboni, A. Shabtai, and Y. Elovici, "Leaking data from enterprise networks using a

compromised smartwatch device," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, ser. SAC '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 741–750. [Online]. Available: https://doi.org/10.1145/3167132.3167214

[53] A. Ostberg, M. Sheik-Nainar, and N. Matic, "Using a mobile device fingerprint sensor as a gestural input device," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 2625–2631. [Online]. Available: https://doi.org/10.1145/2851581.2892419

[54] S. Chen, A. Pande, and P. Mohapatra, "Sensor-assisted facial recognition: An enhanced biometric authentication system for smartphones," in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 109–122. [Online]. Available: https://doi.org/10.1145/2594368.2594373

[55] J. Rose and T. Bourlai, "Deep learning based estimation of facial attributes on challenging mobile phone face datasets," in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ser. ASONAM '19. New York, NY, USA: Association for Computing Machinery, 2020, p. 1120–1127. [Online]. Available: https://doi.org/10.1145/3341161.3343525

[56] S. Lee, W. Choi, and D. H. Lee, "Usable user authentication on a smartwatch using vibration," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 304–319. [Online]. Available: https://doi.org/10.1145/3460120.3484553

[57] J. H. Huh, H. Shin, H. Kim, E. Cheon, Y. Song, C.-H. Lee, and I. Oakley, "Wristacoustic: Through-wrist acoustic response based authentication for smartwatches," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 4, jan 2023. [Online]. Available: https://doi.org/10.1145/3569473

[58] S. Lee, W. Choi, and D. H. Lee, "Usable user authentication on a smartwatch using vibration," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and*

Communications Security, ser. CCS '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 304–319. [Online]. Available: https://doi.org/10.1145/3460120.3484553

[59] C. Cornelius, R. Peterson, J. Skinner, R. Halter, and D. Kotz, "A wearable system that knows who wears it," in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 55–67. [Online]. Available: https://doi.org/10.1145/2594368.2594369

[60] S. Habib, H. Khan, A. Hamilton-Wright, and U. Hengartner, "Revisiting the security of biometric authentication systems against statistical attacks," *ACM Trans. Priv. Secur.*, vol. 26, no. 2, apr 2023. [Online]. Available: https://doi.org/10.1145/3571743

[61] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 82–94. [Online]. Available: https://doi.org/10.1145/2973750.2973764

[62] K. Wu, Q. Yang, B. Yuan, Y. Zou, R. Ruby, and M. Li, "Echowrite: An acoustic-based finger input system without training," *IEEE Transactions on Mobile Computing*, vol. 20, no. 5, pp. 1789–1803, 2021.

[63] A. Turpin, V. Kapitany, J. Radford, D. Rovelli, K. Mitchell, A. Lyons, I. Starshynov, and D. Faccio, "3d imaging from multipath temporal echoes," *Physical Review Letters*, vol. 126, no. 17, Apr 2021. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.126.174301

[64] J. Kim and I. Oakley, "Sonarid: Using sonar to identify fingers on a smartwatch," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: https://doi.org/10.1145/3491102.3501935

[65] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[66] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017.

[67] Y. Chen, T. Ni, W. Xu, and T. Gu, “Swipepass: Acoustic-based second-factor user authentication for smartphones,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 3, sep 2022. [Online]. Available: https://doi.org/10.1145/3550292

[68] J. Kim, J. Park, and I. Oakley, “Sonarauth: Using around device sensing to improve smartwatch behavioral biometrics,” in *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing*, ser. UbiComp/ISWC ’23 Adjunct. New York, NY, USA: Association for Computing Machinery, 2023, p. 83–87. [Online]. Available: https://doi.org/10.1145/3594739.3610696

[69] R. A. Grier, “How high is high? a meta-analysis of nasa-tlx global workload scores,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 59, no. 1, pp. 1727–1731, 2015. [Online]. Available: https://doi.org/10.1177/1541931215591373

Acknowledgement

First and foremost, I greatly appreciate my supervisor, Prof. Ian Oakley. Your guidance, mentorship, and unwavering support have guided me to open new eyes to the HCI research. Working with you during this journey has always been my honor and pride.

I would also like to express my appreciation to the committee members, Prof. Kyungho Lee and Prof. Young-woo Park, for their invaluable feedback, insights, and commitment to my academic growth.

To my Interaction Lab family - DoYoung, Youngeun, Hongmin, Yonghwan, Eunyong, Mingyu, Hyunjae, Mintra, Youryang, Brandon, and Ginger: working alongside each of you has been a remarkable journey. Your friendship, knowledge, and shared experiences have enriched my research and life.

A special mention to my colleagues, Taeyoon, Soobin, Junyoung, Jiyoung, Soohwan, Dongchan, and Jaegyeong. Your friendship and the countless discussions we've had have been a great source of inspiration.

My heartfelt gratitude goes to my family, especially my parents and brother. Their love, belief, and constant encouragement have been a strong support during my Ulsan life. Thank you so much for being there for me.

Lastly, to my partner, Jini (Gyeongjin), your love, patience, and belief in me have been my anchor throughout this journey. This accomplishment would not have been possible without your fantastic Jini effect. I'm looking forward to our new chapter of this journey in Daejeon and Incheon :)