EFRing: Enabling Thumb-to-Index-Finger Microgesture Interaction through Electric Field Sensing Using Single Smart Ring

TAIZHOU CHEN, School of Creative Media, City University of Hong Kong, China
TIANPEI LI, School of Creative Media, City University of Hong Kong, China
XINGYU YANG, School of Creative Media, City University of Hong Kong, China
KENING ZHU*, School of Creative Media, City University of Hong Kong, China
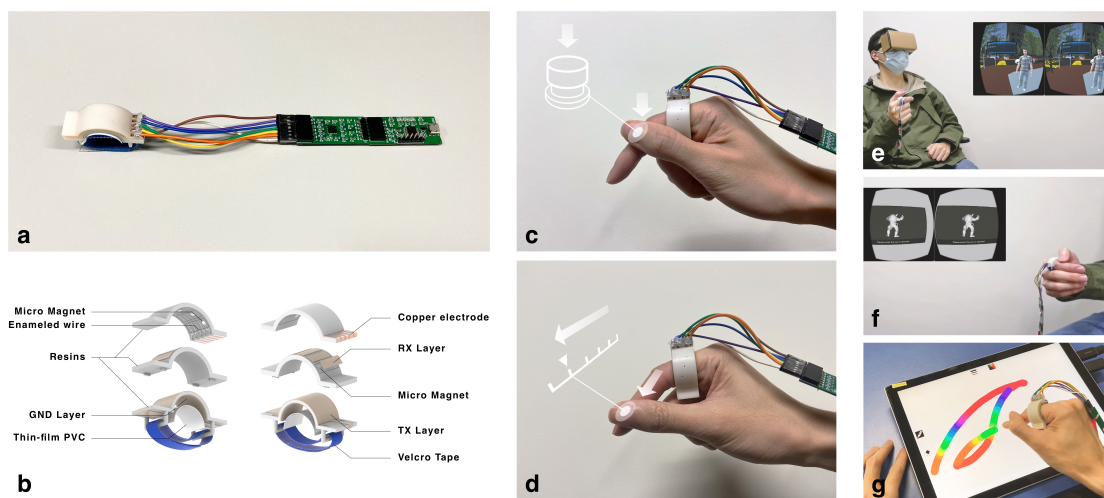
Fig. 1. (a) The EFRing prototype, (b) The multi-layer structure of EFRing, (c) EFRing is able to recognize discrete gestures, such as tapping the thumb on the index finger as a virtual button, (d) EFRing is able to recognize the continuous motion, such as sliding the thumb on the index finger as a virtual slider. (e)(f)(g) show EFRing 's potential applications on Cardboard VR interaction and 2D sketching task.

*Corresponding author. He is also with City University of Hong Kong Shenzhen Research Institute, Shenzhen, China.

Authors' addresses: Taizhou Chen, School of Creative Media, City University of Hong Kong, Hong Kong, China, taizhchen2-c@my.cityu.edu.hk; Tianpei Li, School of Creative Media, City University of Hong Kong, Hong Kong, China, tianpeili2-c@my.cityu.edu.hk; Xingyu Yang, School of Creative Media, City University of Hong Kong, Hong Kong, China, xingyyang8-c@my.cityu.edu.hk; Kening Zhu, School of Creative Media, City University of Hong Kong, Hong Kong, China, keninzhu@cityu.edu.hk.

We present EFRing, an index-finger-worn ring-form device for detecting thumb-to-index-finger (T2I) microgestures through the approach of electric-field (EF) sensing. Based on the signal change induced by the T2I motions, we proposed two machine-learning-based data-processing pipelines: one for recognizing/classifying discrete T2I microgestures, and the other for tracking continuous 1D T2I movements. Our experiments on the EFRing microgesture classification showed an average within-user accuracy of 89.5% and an average cross-user accuracy of 85.2%, for 9 discrete T2I microgestures. For the continuous tracking of 1D T2I movements, our method can achieve the mean-square error of 3.5% for the generic model and 2.3% for the personalized model. Our 1D-Fitts'-Law target-selection study shows that the proposed tracking method with EFRing is intuitive and accurate for real-time usage. Lastly, we proposed and discussed the potential applications for EFRing.

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

Additional Key Words and Phrases: gesture input, sensing technique, machine learning

1 INTRODUCTION

Thumb-to-index-finger (T2I) microgesture interaction [8, 21, 83], such as tapping, rubbing, and circling with the thumb on the index finger, has shown a wide range of benefits for natural [21], efficient [34], and privacy-preserved [8, 83] input. It can be applied to text input [77, 82, 83], mobile interaction [21, 85], drone controlling [84], Internet of Things (IoT) [13], and AR/VR interaction [47]. Thumb gesturing on the index-finger pad is akin to operating a touchable surface (e.g. touch screen) on the index finger. Compared with performing gesture in the mid air [33, 41] or on other large surfaces [11, 45], T2I microgesture interaction could potentially support more light-weight and private interaction with less physical demand.

Attracting increasing research interest in recent years, detecting T2I microgestures is considered to be a challenging problem due to the small range and the occlusion problem of thumb motion [21]. Recent researches have proposed to use different sensors and recognition methods to support such type of inputs, including millimeter-wave radar [25, 48, 76], front-facing passive infrared (PIR) sensors [21], RGB Camera [7], on-thumb motion sensors [47, 49, 87], magnetic sensors [8], and touch-sensitive foils [61, 77, 82, 83]. Among these existing solutions, the finger-worn devices and sensors (e.g., ring-form and nail-mounted devices) [13, 21, 34, 47, 76, 82, 83, 85, 87] are more flexible with higher social acceptance, compared to the glove-like hand-worn devices [61, 77]. While there exists ring-form solutions for detecting thumb gestures [47, 49, 87], they are mostly worn on the thumb. However, thumb-worn rings and accessories could be less common in our daily life, compared to the rings worn on the index fingers [44].

In this paper, we present EFRing, an index-finger-worn ring-form devices for recognizing and tracking T2I microgestures using the electric-field (EF) sensing technique. Our experiments demonstrated that the EF signal could be captured in a high sampling rate (∼200 Hz) and sensitive to the subtle thumb movement without requiring additional instrumentation on the thumb. EFRing detects the thumb movement by emitting an electric field, which would be distorted if a conductive object (e.g., human thumb) approached. The distortion of the electric field could be used for inferring the status of the thumb movement, such as the type of the thumb gesture and the position of the thumb. With EFRing, we design a machine-learning-based recognition pipeline to detect discrete and continuous T2I microgestures. To evaluate the EFRing's feasibility of detecting discrete thumb microgestures, we selected 9 T2I microgestures based on the literature and the user preference. Through our experiments with different classification methods, we showed that the signals captured through EFRing could be classified to recognize the selected T2I microgestures in an average within-user accuracy of 89.5% and an average cross-user accuracy of 85.2%. We further investigated the feasibility of EFRing for tracking continuous 1D T2I motion using a regression-based pipeline. That is, mapping the EFRing signals to the continuous 1D positions

of the thumb while sliding on the index finger. The offline experiment showed a mean-square tracking error of 3.5% for the generic model and 2.3% for the personalized model. We also conducted a 1D-Fitts'-Law user study to evaluate the real time performance and the user experience of the system. The result showed that with EFRing, our participants could perform the 1D-target-selection, achieving a similar performance on 1D target selection on the touch screen in the previous research. The usability questionnaire further shows that the continuous T2I tracking system with EFRing is easy to use with high accuracy. The data set, together with the source code, and the hardware-design documents are available online. [1]

We summarize our contributions as follows:

- We developed EFRing, a ring-form device for capturing the EF signals induced by the TI2 microgestures. Compared to the existing works which either require the devices to be worn on the thumb or need multiple devices on different fingers, we only require one ring worn on the index finger.
- We developed a set of machine-learning-based signal-processing pipelines to recognize discrete T2I microgestures and track continuous T2I motion with EF data recorded by EFRing.
- We evaluated the proposed EFRing hardware and the signal-processing algorithms by thorough experiments and user evaluations.
- We demonstrated a set of proof-of-concept applications that could be supported by EFRing.

2 RELATED WORKS

2.1 Thumb-to-Finger Gesture Recognition

Gesture-based interaction is an important component of the graphical interactive system that supports a natural input. Extensive researches focus on whole-hand tracking and gesture recognition using various types of sensors, such as RGB cameras [69, 75, 80], RGBD camera [41, 58], monochrome camera arrays [29], thermal camera [33], etc.), capacitive sensor array [1, 19], EMG sensors [42, 51, 89], and magnetic field sensors [10]. However, it may be still non-trivial to directly apply these solutions for T2I microgesture detection, as the small-range thumb motion could be occluded or could not arouse large enough sensor signal.

As one of the emerging input methods in recent years, T2I microgesture could provide a light-weight and private interaction paradigm with low physical demand. Previous researches make attempts to enable T2I interaction through different technologies, such as magnetic sensing [8], radio wave [25, 48, 76], infrared sensing [21], acoustic sensing [88], capacitive foils [61, 77, 82, 83], and marker-based AR tracking system [34]. Capturing the thumb movement is not trivial due to the small range and the occlusion of finger motion. One direct solution is utilizing thumb-mounted motion sensors to track the thumb movement [47, 49, 87]. To achieve a similar goal of capturing thumb movement, Chan et al. [8] mounted a magnet on the thumbnail to sense the thumb movement with a Hall sensor attached on the index finger. Instead of mounting sensors on the fingers, the relative movement between the thumb and the index finger could also be captured by an external front-facing device. The Soli project [25, 48] detects the whole-hand gestures through an active radar-sensing system. Utilizing Soli, Wang et al. [76] proposed a CNN+LSTM-based method for recognizing two T2I gestures (i.e., rubbing and sliding) with an average accuracy of 87%. Gong et al. [21] proposed Pyro, a passive sensing system using infrared signal for recognizing 6 T2I microgestures. Their SVM-based classifier shows a within-user accuracy of 84.9% and a cross-user accuracy of 69.0%.

Although the aforementioned works offer solutions for detecting subtle hand/finger gestures in different contexts, there are still some limitations and challenges that worth to be addressed. For example, Soli and Pyro require users to perform their gestures in front of the external sensors, which may restrict the gesture flexibility and the availability of gesture sensing in the mobile context. Directly mounting these sensors on users' bodies may suffer from the sensor signal occlusion problem. The methods that require users wearing glove [61, 77] may

---

[1]https://github.com/taizhouchen/EFRing

suffer from system bulkiness. The finger-worn capacitive sensor foils [82, 83] require direct touching, thus, it may restrict the interaction area on the device surface only. The ring-form sensing device that can be worn on the finger potentially provides a mobile and wearable solution for T2I microgesture interaction. In our work, with EFRing, users wear one ring device on the index finger only, and it supports both discrete T2I microgesture recognition and continuous T2I 1D-motion tracking.

2.2 Ring-Form Interactive Devices

Finger-worn ring-form devices are usually lightweight with high social acceptability [50] to support always-available interaction. Researchers have proposed various types of ring-form wearable devices for both input [7, 39, 47, 60, 79, 87] and output [36, 37, 63, 91].

As the input devices, smart ring-form devices have been used for hand-gesture recognition [7, 78, 79], on-ring touch interaction [5, 13, 70], thumb-movement detection [71, 87], and hand-gesture interaction on the other surfaces [26, 39, 50, 65]. As one pioneering work on finger-worn devices, Masaaki Fukumoto and Yasuhito Suenage [17] proposed a full-time wearable interface named "FingerRing", which senses typing shocks of five fingers with the finger-worn accelerometers. They firstly proposed the concept of "Full-time Wearable Interface" to demonstrate the advantage of small wearable device (e.g., finger ring) that allows always-available interaction without complicating daily life. Wilhelm et al. developed eRing [79] to recognize twelve whole-hand postures using electric-field sensing with a 1-Nearest-Neighbor classifier, achieving an average accuracy of 90%. Perisense [78] classifies eight whole-hand postures in an average accuracy of 98% through electric-field sensing with an IMU.

FingerSound [87] leverages a thumb-mounted ring device consisting of a gyroscope and a contact microphone to detect the thumb gestures performed on the palm. It recognizes 10 uni-stroke digits and 28 Graffiti gestures using K-Nearest-Neighbor (KNN) model. Recently, Liang et al. [47] proposed DualRing, a finger-worn ring-form device to capture the state and the movement of the user's hands and fingers by two motion-sensor-embedded rings wearing on the index finger and the thumb, respectively. Kienzle et al. [40] developed ElectroRing, an index-finger-worn ring that robustly detects the binary T2I pinch by coupling a high-frequency AC circuit. Shi et al. [65] proposed a simple but robust sensing approach to detect if the index finger is contacted with a surface using one motion sensor attached on the index fingernail. Chan et al. [7] developed Cyclopsring to detect full hand gestures through a ring-mounted fisheye-lens RGB camera. Moreover, several researches [2, 5, 13, 32, 70] have explored detecting gestures that are performed directly on the ring device. Darbar et al. [13] developed RingIoT, a smart ring integrated with a motion sensor, a capacitive touch pad, and an IR emitter for Internet-of-Things (IoT) devices control. RingIoT senses the finger pointing direction by the built-in motion sensor and detects the on-ring gesture-based commands using the touch pad. Tsai et al. [70] and Boldu et al. [5] explored detecting five on-ring gestures using a capacitive touch-sensor array.

In our work, EFRing focused on T2I microgrestures, while the existing works studied either large finger gestures (e.g., FingerSound) or whole-hand postures (e.g., eRing, Perisense). Compared to eRing (10Hz) and Perisense (100Hz), EFRing could achieve a higher sampling rate of 200Hz, and require fewer types of sensors than Perisense. FingerSound [87] with the thumb-worn IMU-based ring focused on the large thumb-to-palm gestures, but it may not work for microgesture due to the low sensitivity of IMU. Furthermore, with EFRing , we achieved continuous T2I 1D-motion tracking which was missed in the existing works. For the existing works which required the ring devices to be worn on the thumb which may not be common for ring wearing [44], EFRing offers a solution for the people who prefer the index-finger rings. For those with the ring devices being worn on the index finger, mostly on the proximal phalanx, only on-ring interaction is supported, though the previous surveys show that the finger-pad surfaces are more preferred for microgesture interaction [34]. In addition, direct on-ring interaction could potentially cause discomfort [64]. In our presented work, the user wears the EFRing on the proximal

phalanx of his/her index finger which is a more common and comfortable possible of ring wearing [44], and he/she performs the T2I microgestures on the pad surface of the index finger.

2.3 Sensing with Electric Field

Early in 1995, Zimmerman et al. [92] proposed a taxonomy of electric-field sensing mechanism by categorising the sensing principle into two operating modes: shunt mode and transmit mode. Over the following two decades, many researches [24] have focused on electric-field sensing to support a wide range of human-computer-interaction tasks such as hand-gesture recognition [15, 22, 23, 62], body-posture detection [3, 12], dynamic activity recognition [59], mobile-phone interaction [20], smart-watch interaction [90], and smart furniture [92]. Specifically, J. Rekimoto [62] developed a wrist band incorporated with 7 electrodes and one tilt sensor for hand-gesture detection. Smith et al. [67] proposed a seminal approach that leverage the electric-field sensing mechanism for graphical user interface (GUI) control using mid-air gesture. Grosse-Puppendahl Tobias and Braun Andreas [23] proposed to use an array of capacitive proximity sensors with the combination of frequency- and time-multiplexing algorithms for hand-gesture detection and object recognition. Goc et al. [20] developed a transparent electric-field sensor to support 3D in-air gestures for mobile-phone operation. Zhou et al. [90] proposed AuraSense, an electric-field sensing system for detecting 8 around-watch hand gestures using a SVM classifier, achieving averagely 82.8% accuracy. Recently, Bello et al. [3] developed MoCapaci to sense upper body postures by an in-garment electric-field sensing system incorporated with an CNN-based recognition algorithm with an average accuracy of 86.25% for 20 postures.

EFRing is largely inspired by the aforementioned related works of electric-field sensing for human-computer interaction. We specially focus on recognizing the thumb-to-index finger-microgestures using electric field sensing mechanism, and designing and implementing the solution on the form factor of finger ring.

2.4 Summary

Table 1 summarises seven most related works in the scope of our research. The works with higher accuracy or larger number of gestures either focused on the larger-motion thumb-to-palm or full-hand gestures (i.e., Soli [25, 48], FingerSound [87], eRing [79]), or required more devices to be worn on both the thumb and the index finger (i.e. DualRing [47]). To the best of our knowledge, there are no existing works that support sensing both discrete and continuous T2I gestures with an index-finger-worn ring which EFRing specifically focused on.

3 EFRING: SENSING PRINCIPLE, DESIGN, AND FABRICATION

In this section, we present in details the investigation of the EF-sensing technology around the index finger for detecting T2I microgestures. In our current stage, we mainly focused on the situation of the device being worn on the index finger, due to three main reasons: 1) the index finger was the most common finger for wearing the interactive ring devices in the existing works [74]; 2) the index finger is the closest finger to the thumb, so theoretically the device worn on the index finger could yield better EF signal than that being worn on the other non-thumb fingers; 3) compared to the situation of the device worn on the other non-thumb fingers, it is easier for the thumb to reach the device on the index finger, thus potentially facilitating the on-ring interaction, such as button [2] and touch-pad clicking [5, 70]. We chose EF as the sensing technology mainly due to its potentially high sensitivity, while the other types of sensors (e.g., optical-proximity, ultrasonic, and RF-based sensors) are relatively too bulky to be integrated in a small ring form-factor. Camera could be easily affected by the lighting condition. Based on the technical investigation of EF sensing, we will discuss the design and the making of EFRing.

Table 1. Summary of comparison with previous works

|	Sensors	Gesture Types	Gesture Number	Accuracy	Form-Factor
Pyro [21]	PIR sensor	T2I gestures	6	With-in user: 84.9% Cross-user: 69.0%	Non-wearable
Soli [25, 48]	Radar	Full-hand gestures	11	Accuracy:87.2%	Non-wearable
Thumb-in-motion [5]	Touch film	On-ring gestures	5	F1 score: 86.9% Recall: 89%	Index-finger ring
DualRing [47]	IMU +electrode	T2I gestures	4	Accuracy:99.4%	Index-finer ring and thumb ring
FingerSound [87]	IMU +Microphone	Thumb-to-palm gestures	28	Accuracy:92.5%	Thumb ring
eRing [79]	EF	Full-hand postures	6	Accuracy:97%	Index-finger ring
Perisense [78]	EF+IMU	Full-hand postures	8	Accuracy:88%	Index-finger ring
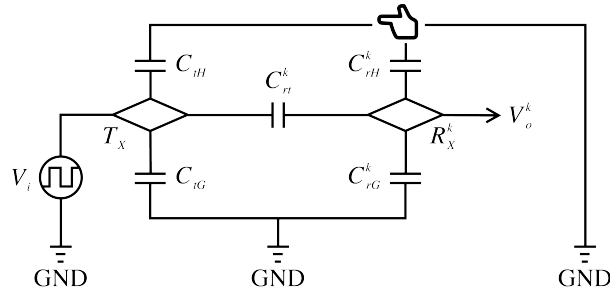
Fig. 2. The illustration of a circuit model for electric-field sensing system.

3.1 Sensing Principle

In general, electric field sensing is based on a principle that any charge-carrying surface (e.g., electrode) can generate an electric field, which will be distorted if any conductive object (e.g., finger) approaching [73]. As shown in figure 2, an electric-field sensing system usually consists of three layers: the receiver layer (Rx), the transmitter layer (Tx), and the ground layer (GND). A transmitter layer is usually driven by a low-voltage alternating-voltage (AC) signal (115kHz, 3Vpp square wave for EFRing), and the receiver layer will detect the transmitted AC signal through the capacitive path between the transmitter electrode and the receiver electrode. When a person's hand or finger intrudes the electrical field, the low-frequency energy will be capacitively coupled into the human body. Since the human body is much larger than the size of the electrode, the portion of the human body that is out of the field serves as another charge reservoir [92]. In other words, the coupling of the human body would cause the change of the capacitance of the whole system. Such changes could be processed to track the body movements and classify the movement patterns.

Based on the aforementioned principle, we implemented a multiple-receiver electric-field sensing system for EFRing to enrich the sensing resolution for detecting the subtle finger movement. EFRing detects the thumb-to-index finger movement by generating an electric field around the ring-worn position around the index finger. If

the relative movements between the thumb and the index finger occurred, the system capacitance will change, and the output voltages from the receiver electrodes will also change correspondingly. Given an input voltage as V_i to the Tx, the output voltage V_o^k from the kth Rx is formulated by [57]:

$$V_o^k = V_i \times \frac{C_{rt}^k}{C_{rt}^k + C_{rG}^k + C_{rH}^k} \tag{1}$$

where C_{rt}^k is the capacitance between the kth Rx and Tx, C_{rG}^k is the capacitance of the kth Rx to GND, C_{rH}^k is the capacitance between the kth Rx and human finger. In practice, we normally implement the input voltage V_i as a square wave AC signal generated by an oscillator. For more details, please refer to the IC data sheet [57].

The capacitance C (in Farads) between two electrodes is given by

$$C = \varepsilon_r \varepsilon_0 \frac{A}{d} \tag{2}$$

where A is the overlapping area (in square meters) between the two electrodes, d is the relative distance (in meters) between the electrodes, ε_r and ε_0 are relative permittivity and vacuum permittivity respectively. ε_0 is normally treated as a constant value according to Glauser et al. [18]. ε_r is usually affected by the material permittivity and the vacuum permittivity. Therefore, in our case, ε_r is also a constant value since all the insulation layers on EFRing are fabricated with the same type of resin in our implementation. Combining Equation 1 and 2, we can see the generated electric field could be affected by A and d mainly, which means that the placements of the transmitter and the receiver electrodes would affect the sensing range of the system. In the actual hardware implementation, the relative positions between the Rx(s), the Tx, and GND usually don't change. Thus, given equation 2 and equation 1, we have

$$V_o^k \propto \frac{d_{rH}^k}{A_{rH}^k} \tag{3}$$

indicating the output voltage V_o^k from the receivers Rx(s) changes proportionally to the distance between the finger and Rx(s), and changes inversely proportional to the overlapping area between them. Therefore, it is reasonable to hypothesize that the change of V_o^k may potentially reflect the movement pattern of the thumb which could indicate the type of the T2I gestures. We further verify our hypothesis with a serial of experiments in Section 5 and Section 6.

3.2 Designing and Making EFRing

For EFRing, we aim to build a lightweight ring device that is comfortable to wear, and the ring needed to be flexible for fitting different finger sizes. Figure 1b illustrates the decomposition of our final design of EFRing. We design a multi-layer structure to construct the antenna. The key components of EFRing are the Rx, the Tx, and the GND layers, which are made of thin-film conductive fabric and produced by the laser cutter. The top three insulation layers are fabricated using a SLA 3D printer, and the bottom one is made by thin-film polyvinyl chloride (PVC). For the Rx antenna layout, we placed five thin-film antenna to cover a sensing angle range of 120° (Fig. 3) in the receiver layer, to maximize the range of the received signals. Noted that the current antenna layout is in 1-dimensional. We didn't use the 2D antenna layout as it would reduce the size of the individual antenna and shorten the distance between two adjacent antennas. Smaller antenna and more dense antenna layout may lead to the interference among the antennas, and would potentially affect the sensing capability of the system [31]. While the antenna design could be further optimized, it is out of our scope in the current stage.

One challenge of constructing EFRing is to properly connect Tx and Rx(s) into the circuit meanwhile preventing short circuit between them within a small compartment. The connection should also be unobtrusive to reduce the

external interference over the electric field. To this end, we use five 0.1mm-thin enameled wires to connect the five Rx(s) to the copper electrodes. For each enameled wire, we use two micro magnets to ensure its proper contact with the antenna. Our pilot test shows that the intrusion of micro magnets would cause a constant interference on the generated electric field in EFRing. Such constant interference could be eliminated by calculating the derivative of the temporal signal, as proved by previous works [46, 87].
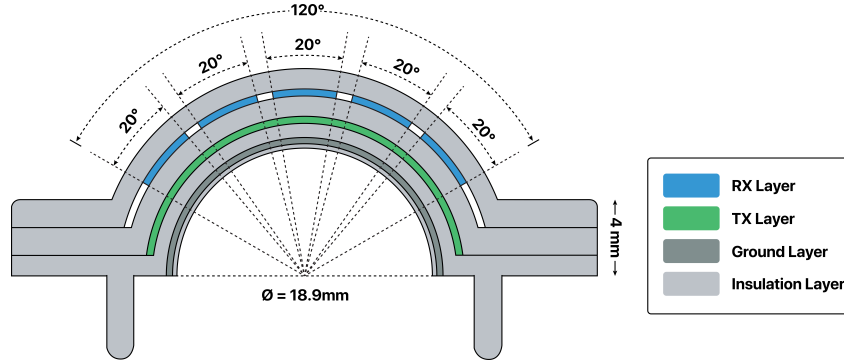

Fig. 3. Cutaway view of EFRing. The Rx(s) are covering a sensing angle range of $120°$, while each of them is covering a range of $20°$.

To capture the electric-field signal, we customize a PCB board based on Microchip MGC3130 which is connected to the antennas on EFRing. The circuit was referred to the reference schematic as suggested by the MGC3130 data sheet [57]. We configure the transmitter to generate an electric field using a 115kHz, 3V square wave. The chip also monitors each of the Rx electrodes to measure the voltage attenuation at a sample rate of 190Hz. The measurements were output through the I^2C protocol to a USB bridge [56] to communicate with a PC.

4 PILOT EXPERIMENT: OPTIMIZING THE DEVICE ORIENTATION

Ring is often considered to be omnidirectional while being worn. To this end, we first conducted a pilot experiment to investigate the signal quality under different wearing orientations. Equation 2 shows that A and d are two key factors that affect the quality of the electric-field signal. Figure 4 shows illustrations under three different wearing orientations: $0°$, $45°$, and $90°$, resulting in different antenna directions (the light blue regions in Figure 4). The antenna orientation affects the change of A and d between the thumb and the antenna while users performing T2I microgestures. Therefore, there is a need for finding a suitable wearing angle for EFRing to capture sufficient information of the thumb movement.

To optimize the quality of the EF signal for EFRing, the wearing orientation needs to satisfy two criteria: 1) The signal profiles are distinguishable across a variety of T2I microgestures, and 2) the signal profile for each unique gesture is stable across different users.

As our T2I gestures were performed on the YZ-plane (Fig. 4), we could interpret that, according to Equation. 4, if we move the thumb along the Y axis, the projection A on the antennas (light blue area) would change. If we move along the Z axis, the distance d between the thumb and the antennas would change. Compared with $0°$ and $45°$, the antenna area under $90°$ covers the YZ-plane the most, allowing more antennas to catch the changes of A and d which lead to the change of output voltage induced by the thumb movement. Therefore, we anticipated that $90°$ is the optimal wearing orientations for our task. To further verify this anticipation/hypothesis on the angle setting, we adopted a data-driven approach to investigate the signal quality under different ring-wearing
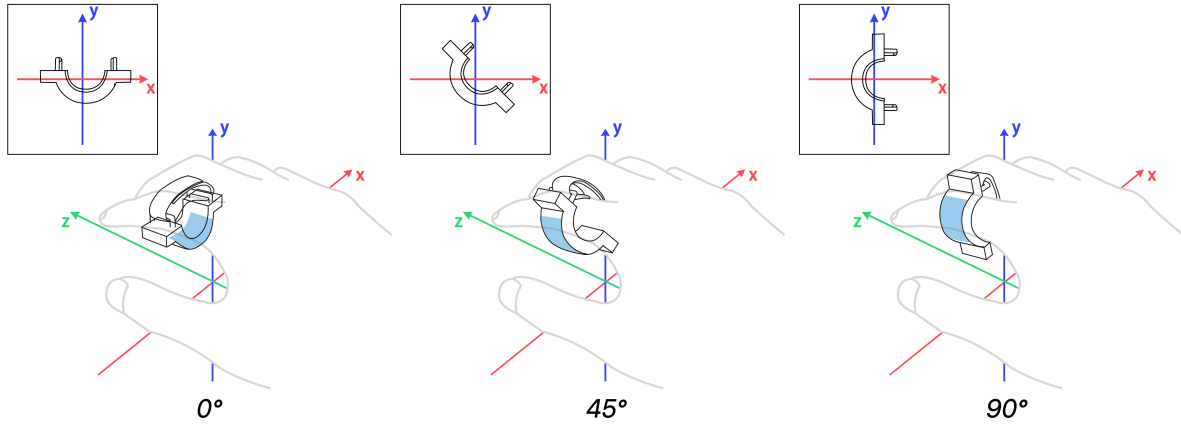
Fig. 4. Three chosen device orientations for our pilot experiment. The light blue areas indicate the placement of the Rx(s) antennas.

orientations. Specifically, we collected the antenna signals of different users wearing EFRing and performing a set of selected T2I microgestures, and analyzed the collected signals in high dimensional space.

4.1 Data Collection

We recruited 3 participants (2 males, 1 female, mean age = 24.67, SD = 2.08) from a local university through words of mouth. All of them are right-handed without any prior experience on smart ring. The participants were instructed to wear EFRing on the proximal phalanx of their index fingers in three different angles. Under each wearing angle, the participants performed 10 pre-selected thumb-to-index finger gestures (see Section. 5.1 for details). During the recording of each gesture, the participants were told to perform the gestures within 1.2 seconds with a rolling animation shown on the screen as the time indicator. For each gesture under each wearing angle, we recorded 100 samples from each participant. Participants were required to take off the ring and take it on again after every 10 recordings to increase the data variance. The recording order of angles was counter-balanced across three participants, and the order of the gestures under each angle was randomly shuffled. To record the data, we connected EFRing to a laptop (Window 10, Intel Core i5, 16GB RAM) through USB connection in a sample rate of 190Hz. We developed the data-recording interface using Qt5.6 in C++. After removing the corrupted data due to software and hardware error, we recorded 6993 gesture samples in total, with 2112 for $0°$, 2304 for $45°$, and 2577 for $90°$.

4.2 Data Analysis and Results

Based on the criteria of the optimal wearing angle described in Section 4.1, we define the signals of each gesture as one cluster, and focus on the between-clusters dispersion and inter-cluster dispersion of the signals for each angle. Specifically, for a specific wearing angle, a better gesture distinguishability leads to a larger between-clusters dispersion in the latent space, while a good feature stability across different users precipitates a small inter-cluster dispersion in latent space. Specifically, we calculated *Calinski-Harabasz Index* [6] to indicate the between-clusters dispersion and inter-cluster dispersion of the microgesture signals. Given a between-clusters sum of square SS_{BG} and a inter-cluster sum of square SS_{WG}, the *Calinski-Harabasz Index CH* is given as:

$$CH = \frac{SS_{BG}}{K-1} \bigg/ \frac{SS_{WG}}{N-K} \qquad (4)$$

where K is the total number of clusters, and N is the total number of samples. With the ground truth cluster label for each gesture sample in the latent space feature point, SS_{BG} and SS_{WG} can be calculated as:

$$SS_{BG} = \sum_{k=1}^{K} n_k \|c_k - c\|^2 \qquad (5)$$

$$SS_{WG} = \sum_{k=1}^{K} \sum_{x \in C_k} \|x - c_k\|^2 \qquad (6)$$

where k represents the kth cluster, C_k is the set of points in k, c_k is the cluster center of k, and n_k is the number of points in k. c is the center of all sample in the space. The value of *Calinski-Harabasz Index* is proportional to the between-clusters dispersion and inversely proportional to the inter-cluster dispersion, which could indicate the signal quality under a specific wearing orientation.
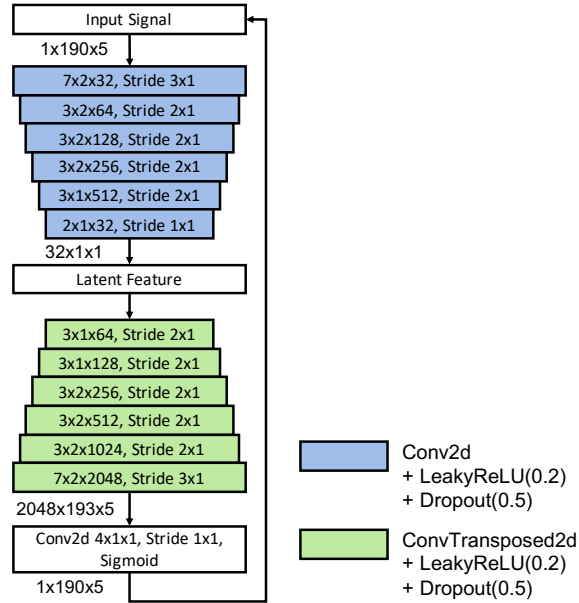

Fig. 5. The illustration of our auto-encoder model. The model was trained to reconstruct the input signal. After training, we extracted the latent feature as a low-dimensional representation of the input signal for further analysis.

The *CH* value could be computed using different forms of cluster features, such as the hand-crafted statistical feature vectors and the latent feature vectors extracted by auto encoders. As the hand-crafted statistical features may be directly affected by noise, we adopted the auto-encoder approach for feature extraction. Specifically, we train a CNN-based encoder-decoder model for self-reconstructing the collected raw signals in time series. The latent feature vector was used as a low-dimensional representation of the signal for *CH* calculation.

The auto-encoder-decoder model (Fig. 5) consists of 6 2D convolutional layer for the encoder and 6 2D transposed convolutional layer for the decoder. Here we consider the 2D convolutional structure over the 1D

	(a) Tap	(b) Double Tap	(c) Swipe Left	(d) Swipe Right	(e) Swipe Up
Ease	7.00	7.00	6.75	6.81	5.88
SocialAcc	6.94	6.94	6.75	6.81	6.44
Fatigue	6.88	6.75	6.81	6.75	6.19

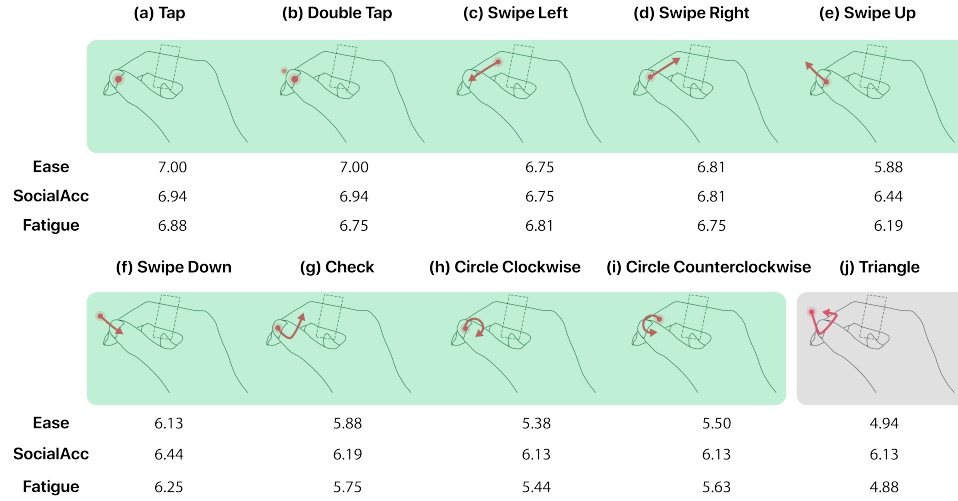|	(f) Swipe Down	(g) Check	(h) Circle Clockwise	(i) Circle Counterclockwise	(j) Triangle
Ease	6.13	5.88	5.38	5.50	4.94
SocialAcc	6.44	6.19	6.13	6.13	6.13
Fatigue	6.25	5.75	5.44	5.63	4.88

Fig. 6. Illustration of the initial gesture set and their corresponding user rating. **Ease** is a short of *ease to perform*, while **SocialAcc** is a short *Social acceptance.* The rating score of **Fatigue** is subtracted by 7. Gestures with the light green background indicate the final gesture set.

model, as the 2D model could extract the channel-wise features which could be potentially helpful. The model takes the original 1-second time-series signal as input, and regenerate the signal sequence as output. We calculated the mean-square error as the reconstruction loss during training. The size of the latent-space feature vector is 32. The model was trained on a desktop PC with dual RTX 2080 NVIDIA GPU, 32GB RAM, and one Intel i7-8700 CPU. We used an Adam optimiser [43] ($\beta1 = 0.9$, $\beta2 = 0.999$) with the learning rate of $1e - 5$ for model optimization. Our implementation was based on PyTorch 1.10 with Python 3.8. For each auto-encoder model, we train for 128 epoch in a batch size of 16. The result suggests the *CH* value of $90°$ (55.04) outperforms the other two angles ($45°$: 36.13, and $0°$: 24.12). As a result, we finalized $90°$ as the device-wearing orientation of EFRing for the following experiments.

5 DISCRETE SUBTLE THUMB-TO-INDEX FINGER GESTURES

In this section, we investigated the feasibility of EFRing for recognizing discrete T2I microgestures. Here we define the discrete gesture as a gesture that is completed with a sequence of motions (e.g., drawing a circle) and could be mapped to a specific command/event. We first selected a set of T2I microgestures with the consideration of user preference. Then, we proposed and experimented a machine-learning-based signal-processing pipeline for gesture classification.

5.1 Gesture Selection

We first summarized an initial gesture set from existing literatures on thumb-to-fingers microgesture interaction [21, 74]. Specifically, we pick 6 unistroke thumb-to-index finger gestures (*Tap*, *Double Tap*, *Swipe Up*, *Swipe Down*, *Swipe Left*, *Swipe Right*) from a recent work on ring-based gesture literature review [74], and 3 gestures (*Triangle*, *Check*, *Counterclockwise Circle*) from [21]. To maintain the design consistency, we add one more gesture(*Clockwise Circle*) in to the gesture set. As a result, we built up the initial gesture set with 10 T2I microgestures as shown in Fig. 6. Following the previous works on gesture design and selection [11, 81], we then conducted a focus-group

survey on the user preference toward 10 gestures in our initial gesture set, to narrow down the focus of T2I microgesture set.

5.1.1 Participants and Procedure. Through the word of mouth, we recruited 16 participants (8 males and 8 females) with the average age of 25.8 years old (SD=2.40). All of them are right-handed, without any experience of smart ring. The experimenter began the survey by asking the participant to fill a pre-questionnaire with demographic information, and the participants wore EFRing on the proximal phalanx of their index fingers. In addition, the participants were introduced to the whole set of the gestures before the survey formally started.

During the survey, the experimenter randomly showed a gesture illustration (Fig. 6) to the participant and ask them to perform the gesture several times until they felt they can perform the gesture correctly without looking at their fingers. We then asked the participants to rate each gesture according to three criteria along a 7-point Likert-scale (1: strongly disagree to 7: strongly agree):
- *Ease to perform*: "It is easy to perform this gesture precisely."
- *Social acceptance*: "The gesture can be performed without social concern."
- *Fatigue*: "The gesture makes me tired."

The participants were also invited to discuss the rationale behind their ratings, and the verbal discussion was recorded.

5.1.2 Result. A multi-factorial repeated-measures ANOVA was performed on the ratings of ease to perform, social acceptance, and fatigue. The results showed a significant effect of the gesture type on the ratings of ease to perform ($F(9,135) = 13.87$, $p < 0.005$, $\eta_p^2 = 0.48$), social acceptance ($F(9,135) = 6.01$, $p < 0.005$, $\eta_p^2 = 0.286$), and fatigue ($F(9,135) = 11.82$, $p < 0.005$, $\eta_p^2 = 0.441$). For each criteria under each gesture, we calculate the average score across different user. Fig. 6 shows the descriptive results of the average ratings for each gesture. We selected the gestures whose three ratings were all above 5 as suggested by previous works on preference-based gesture selection [11, 81]. As a result, this selection process eliminated only one T2I microgesture: *Triangle*, leaving 9 gestures as highlighted in light green background in Fig. 6.

5.2 Machine-Learning-based EFRing Gesture Classification

After finalizing the gesture set, we experimented with the feasibility of EFRing for recognizing the gestures through a machine-learning-based approach. We first construct a data set of the EFRing signals for the selected gestures. Then we propose a deep-learning architecture to process multi-channel EFRing signals for classifying the selected discrete gestures.

5.2.1 Data Acquisition. We recruited 16 participants (8 males, 8 females, mean age = 25.50, SD = 2.98) from a local university for data collection. All of them are right-hand dominant without any prior experience of smart ring. The setup of the data acquisition for the selected gestures was similar to the pilot experiment (Section 4.1). Additionally, we divided the data-collection process into two parts for each participant. During part one, participants were asked to repeat each gesture 40 times until he/she finish all gestures in the random order. We treated the part-one data as our training data. We also asked him/her to take off the ring and take it on again after every 10 times of recording, to increase the data variance. In part two, participants were asked to repeat each gesture 10 times, and the gestures were shuffled randomly. We treated the data recorded in part two as testing data. During the data collection, We did not strictly control the participants' ring-wearing angles. Instead, we softly told the participants the desired wearing angle of 90°, and they wore the ring based on their own perception. This was to further ensure the variety of the data. Similar to Chen et al. [11], during the recording of each gesture, the participant first saw a 3-second count-down on the screen, followed by a 1-second decreasing circular progress bar for performing the gesture. The whole experiment took about 40 minutes for one participant. In the end, two users' data were disgarded because of the accidental hardware failure (i.e. the unexpected disconnection between

the enameled wires and the receiver antenna). As a result, we collected 6095 valid gesture samples, consist of 4832 training sample and 1263 testing sample.

5.2.2 Data Representation and Processing. We denote one measurement frame from EFRing at timestamp t as $x_t = [V_t^1, V_t^2, V_t^3, V_t^4, V_t^5]^T$, where V_t^n represent the output voltage measured from the nth channel (i.e Rx) at t. We denote a gesture sample containing N measurements as $X = \{x_t\}_{t \in \{0,...,N-1\}} \in \mathbb{R}^{(5 \times N)}$. In our experiment, we set N to 190, which is 1 second under the sample rate of 190Hz.

For every gesture sample in our data set, we first removed 19 frames (0.1 seconds) from the head and the tail respectively to eliminate the noisy measurement occurred while activating and terminating the recording program. We then applied the operation *Fourier* re-sampling on every channel to make sure the sample length is 190 after removing the head and the tail. Lastly, we applied a low-pass filter with the cut-off frequency of 80Hz to remove the potential high-frequency random noise signal generated by the circuit. The value of 80Hz for the cut-off frequency was determined as it achieved the highest SNR (Fig. 7) in our empirical experiments where we tested various cut-off frequencies ranging from 50Hz to 94Hz.
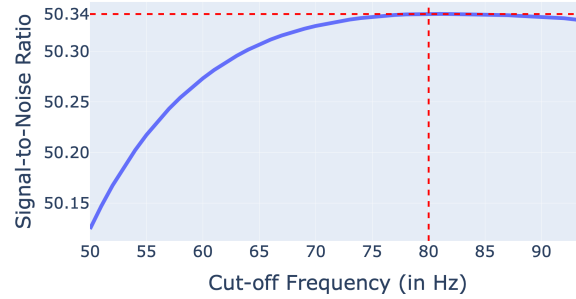

Fig. 7. The average signal-to-noise ratio (SNR) on our data set after applying the low-pass filter with different cut-off frequencies ranging from 50Hz to 94Hz. The x axis is the cut-off frequencies, and the y axis is the SNR after low-pass filtering. The best cut-off frequency was yielded by the cut-off frequency of 80Hz, with the highest SNR of 50.34.

5.2.3 Feature Extraction. For the gesture-classification task using EFRing, we observed a spatial correlation between each pair of Rx channels that could be potentially helpful for the classification-model training and testing. We encode such relationships by pre-computing the pairwise gradient between any two Rx channels. For the pair of the ith and the jth channels (i, j) at time stamp t, we compute their channel-wise gradient as:

$$\Delta V_{(i,j)}^t = \frac{V_i^t - V_j^t}{|i - j|} \tag{7}$$

All the channel pairs (i, j) in a gesture sample X are denoted with P_c. We compute $\Delta V_{(i,j)}^t$ for every pair in P_c for every time stamp to form a new feature set: $dC(X) = \{\Delta V_{(i,j)}^t | t \in \{0, ..., N-1\}, (i, j) \in P_c\} \in \mathbb{R}^{(10 \times N)}$. Apart from computing channel-wise gradient, we also take the signal's temporal gradient: $dT(X) \in \mathbb{R}^{(5 \times N)}$ as the previous gesture-classification works [46, 87] did. We combine the channel-wise gradient and the temporal gradient as our final feature set: $dC(X) \cup dT(X), \in \mathbb{R}^{(15 \times N)}$. We then conducted a series of training and testing experiments for selecting the proper feature and classification model.

5.2.4 Experiments on T2I Microgesture Classification. In this experiment, we first compared a set of machine-learning architectures for classifying the selected discrete gestures using different feature combinations. The possible feature combinations are the channel-wise gradient only (C), the temporal gradient only (T), and the combination of the channel-wise and the temporal gradients (T+C).

Experimental Conditions.

- *Data Usage.* To examine the generalizability of the gesture classification using EFRing, we adopted two data-splitting schemes: 1) We used the data from all 14 users as within-user evaluation, with the split of 3:1:1 for the training, the validation, and the testing data respectively; 2) We trained the models on the data from 11 users and tested them on three left-out users as a leave-three-user-out evaluation. The three left-out users are selected based on their signal quality. Specifically, we calculated the average signal-to-noise ratio (SNR) for each user's data and picked 3 users whose SNRs are three lowest as three leave-three-out testing dataset. During the leave-three-out experiment, we used a random 30% subset of the training data as the validation data.

- *Selection of Classification Models.* We experimented and compared different machine-learning models for discrete T2I microgesture classification, with the EFRing signal features in both the time and the frequency domains. For the signal features in time domain, we experimented three models: Support Vector Machine (SVM), Multilayer Perceptron (MLP, 4 layers, 1024 units per layer), and Convolutional Neural Network (same as the encoder structure described in Section 4.2). To investigate the signal in frequency domain, we applied the process of Short-time Fourier transform (STFT) on each feature channel. For example, under the feature extraction scheme T+C, the STFT process will result in a 15-channel image tensor for each gesture, with each channel is a 2D spectrogram image. Such data structure could be processed and classified using existing widely-used CNN models, such as VGG [66], ResNet [30], DenseNet [35], and Visual Transformer (ViT) [14].

Training. To increase the model's generalizability and avoid over-fitting, we adopted a data-augmentation scheme by rolling each gesture sample along time axis by a random offset. We set the data-augmentation probability to 0.5 for every sample during training. We also employed the label-smoothing technique [68] with the smoothing parameter $\epsilon = 0.1$ during training.

We trained all the chosen models on a desktop PC with dual RTX 2080 NVIDIA GPU, 32GB RAM, and one Intel i7-8700 CPU. We used an Adam optimiser [43] ($\beta1 = 0.9, \beta2 = 0.999$) with the learning rate of $1e-4$ for model optimization. We decayed the learning rate by a factor of 0.1 if the validation loss is not decreasing for at lease 5 epochs. For each model, we trained for 128 epoch in a batch size of 16. Our implementation was based on PyTorch 1.10 with Python 3.8.

Results. Table 2 summarizes the testing accuracy of each classification model with different data feature combinations. The result shows that using the combination of the channel-wise and the temporal gradients (T+C), DenseNet achieved the highest accuracy for the within-user testing of 89.5%, while ViT performs the best on the leave-three-users-out testing, with an average accuracy of 85.2%.

Overall, transferring the time-series signal to the frequency domain achieved a higher performance compared to directly using the signal in the time domain. This could be because the 2D data representation of the signal in the frequency domain allows the usage of deep neural networks for hidden feature extraction. The results also show that in general, using the T+C data leads to a better performance than using T or C only. This suggests the importance of both the channel-wise and the temporal features for T2I microgesture classification on EFRing.

Although the performance of ViT on the within-user data set is inferior to DenseNet ($\tilde{0}.3\%$ lower), it shows a larger advantage on classifying the left-three-users-out testing data ($\tilde{0}.8\%$ larger) which could be less biased. To this end, we chose ViT for further analysis and experiments. Table 3 shows the confusion matrix of the leave-three-users-out cross-validation on ViT. The *Check* (g) gesture is the most challenging one (66%). The ViT

Table 2. Testing accuracy of each gesture with different models. This table shows the accuracy on different model for both within-user evaluation (W) and leave-3-user-out evaluation (L) under different feature extraction scheme: temporal gradient (T), channel-wise gradient (C), and mix gradient (T+C). The upper part of the table shows the models train with feature in the time domain. The lower part of the table is focusing on the evaluation in the frequency domain.

Model	Ftr.	Average W	Average L	(a) W	(a) L	(b) W	(b) L	(c) W	(c) L	(d) W	(d) L	(e) W	(e) L	(f) W	(f) L	(g) W	(g) L	(h) W	(h) L	(i) W	(i) L
										Time Domain											
SVM	T	52.6%	47.8%	64.4%	64.1%	48.7%	57.7%	75.3%	88.0%	49.4%	34.3%	62.2%	29.4%	41.0%	45.5%	33.7%	29.0%	45.3%	46.8%	38.7%	37.5%
	C	67.3%	51.5%	66.5%	60.0%	52.1%	40.4%	86.5%	100.0%	62.3%	41.9%	74.8%	48.6%	68.6%	58.8%	74.2%	51.4%	60.9%	31.3%	65.8%	51.1%
	T+C	75.8%	64.8%	82.4%	82.6%	58.9%	62.2%	87.6%	100.0%	69.9%	53.6%	84.0%	56.8%	78.5%	68.8%	75.6%	50.0%	77.2%	73.3%	77.6%	61.0%
MLP	T	59.1%	51.9%	66.5%	70.0%	55.6%	60.7%	77.8%	84.0%	58.5%	58.3%	67.4%	41.7%	75.9%	37.5%	38.1%	34.0%	52.0%	55.8%	50.0%	34.2%
	C	72.0%	63.3%	71.5%	90.0%	69.3%	64.0%	88.7%	100.0%	64.3%	48.0%	71.9%	45.1%	76.4%	65.0%	79.1%	53.7%	61.5%	69.7%	66.9%	67.7%
	T+C	80.3%	74.1%	88.3%	100.0%	78.0%	85.7%	86.5%	100.0%	72.8%	68.0%	78.2%	60.5%	81.8%	76.2%	75.6%	52.5%	81.3%	87.1%	79.6%	63.9%
CNN	T	75.9%	68.9%	83.9%	95.8%	98.5%	100.0%	82.0%	91.3%	64.5%	44.4%	77.1%	61.1%	59.5%	65.2%	63.2%	37.5%	76.6%	68.3%	81.0%	77.8%
	C	82.7%	78.5%	83.0%	87.1%	94.9%	100.0%	86.7%	100.0%	79.9%	75.0%	76.3%	54.2%	68.8%	80.7%	83.7%	68.6%	83.8%	80.7%	92.3%	85.0%
	T+C	86.1%	78.2%	90.3%	90.3%	95.8%	100.0%	84.2%	90.9%	79.4%	66.7%	80.3%	70.0%	74.2%	76.9%	88.1%	61.3%	91.6%	77.8%	92.3%	78.8%
										Frequency Domain											
VGG	T	82.7%	75.6%	100.0%	96.4%	100.0%	96.6%	82.5%	95.5%	72.3%	72.2%	75.7%	59.5%	69.9%	68.2%	75.0%	55.0%	80.8%	68.2%	91.6%	92.0%
	C	85.1%	82.2%	92.9%	96.6%	98.5%	100.0%	92.0%	100.0%	71.1%	62.8%	80.9%	73.3%	73.2%	81.5%	87.0%	66.7%	86.0%	87.9%	90.2%	83.3%
	T+C	87.3%	78.9%	94.7%	96.7%	100.0%	100.0%	90.1%	100.0%	75.8%	56.3%	90.6%	70.3%	71.8%	69.6%	89.3%	48.6%	89.4%	85.7%	90.6%	96.3%
ResNet	T	80.3%	74.4%	92.3%	100.0%	100.0%	100.0%	88.9%	95.5%	68.3%	50.0%	75.7%	69.0%	64.2%	73.7%	73.5%	46.5%	82.0%	74.4%	82.5%	78.1%
	C	86.1%	76.7%	97.0%	100.0%	98.6%	100.0%	90.2%	92.3%	77.1%	47.8%	82.2%	71.4%	72.0%	72.7%	84.5%	66.7%	85.6%	84.9%	90.6%	71.0%
	T+C	85.7%	76.3%	96.3%	100.0%	100.0%	100.0%	91.0%	100.0%	76.3%	53.5%	78.0%	74.1%	71.4%	70.4%	81.0%	53.9%	90.5%	81.1%	91.6%	70.6%
DenseNet	T	84.3%	77.0%	97.0%	96.4%	97.9%	96.8%	88.9%	96.0%	72.0%	59.1%	84.0%	78.8%	66.7%	65.0%	82.8%	56.8%	87.9%	69.8%	87.8%	77.4%
	C	86.0%	78.9%	98.5%	100.0%	100.0%	100.0%	91.0%	100.0%	72.2%	52.8%	81.1%	71.1%	71.0%	64.5%	86.0%	72.7%	85.5%	87.5%	95.2%	76.5%
	T+C	**89.5%**	84.4%	97.8%	100.0%	100.0%	96.9%	92.3%	100.0%	78.0%	62.5%	81.7%	78.8%	80.4%	75.9%	94.2%	83.3%	92.7%	79.0%	93.4%	89.7%
ViT	T	87.2%	77.0%	97.1%	90.6%	100.0%	100.0%	92.1%	91.3%	74.5%	73.9%	87.3%	83.3%	74.5%	60.7%	81.1%	62.2%	89.1%	65.2%	91.8%	80.7%
	C	87.5%	81.9%	91.8%	96.4%	94.6%	100.0%	90.9%	100.0%	81.2%	75.0%	83.2%	73.5%	78.4%	88.9%	90.3%	63.4%	87.5%	93.8%	91.2%	64.7%
	T+C	89.2%	**85.2%**	95.0%	100.0%	98.6%	100.0%	90.2%	100.0%	81.3%	83.9%	91.5%	87.1%	78.4%	71.0%	85.2%	65.8%	93.7%	85.7%	89.3%	85.2%

model tends to confuse *Check* (g) and *Swipe Right* (d) as both gestures involve the thumb motion towards the right side. The confusion matrix also shows that the model is able to distinguish between the symmetric gestures (e.g. *Swipe Left* (c) and *Swipe Right* (d), *Swipe Up* (e) and *Swipe Down* (f), and *Circle Clockwise* (h) and *Circle Counter-Clockwise* (i)). As for the simple gestures like *Tap* (a) and *Double Tap* (b), the classification performance remains consistently high across different users.

Transfer Learning for Personalized Model. In the real-world scenario, transfer learning has been widely used to adopt a general classification model to a personalized model using a small amount of user-specific data. Therefore, we also experimented the transfer-learning process on the trained ViT models with a small amount of data from the three left-out users. Specifically, we chose a random 20% subset of training data (8 gesture samples per label per person) from these three users. We adopted a data-augmentation scheme by rolling each gesture sample along the time axis by a random offset with a probability of 0.5. The model was tested on three left-out-users' testing set every 5 epochs. The results (Figure 8) show that the average testing accuracy increases from 85.2% to 87.4% after 35 epoch of training.

Table 3. Confusion matrix of ViT from leave-3-user-out cross-validation. The horizontal axes indicates the prediction while the vertical axes annotates the ground truth label. Darker cell-shade indicates higher accuracy.

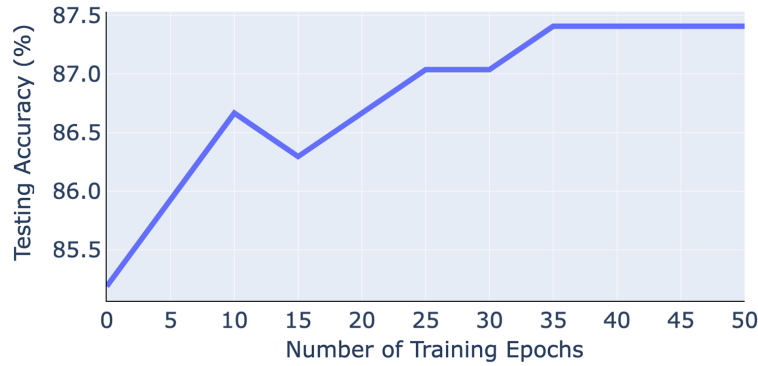| GT	Prediction								
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
(a)	100%	0%	0%	0%	0%	0%	0%	0%	0%
(b)	0%	100%	0%	0%	0%	0%	0%	0%	0%
(c)	0%	0%	100%	0%	0%	0%	0%	0%	0%
(d)	0%	0%	0%	84%	0%	6%	10%	0%	0%
(e)	0%	0%	10%	0%	87%	3%	0%	0%	0%
(f)	3%	0%	6%	3%	3%	71%	3%	0%	10%
(g)	0%	0%	0%	5%	5%	13%	66%	0%	11%
(h)	6%	3%	3%	0%	0%	0%	3%	86%	0%
(i)	0%	4%	11%	0%	0%	0%	0%	0%	85%

Fig. 8. Testing accuracy on the personalized model through transfer learning. The horizontal axis is the number of the training epoch, and the vertical axis is the accuracy on classifying the testing set using the personalized model.

Ablation Study on the Number of Receiver Antennas. To study the effectiveness of the multi-receiver EF system in EFRing, we ran an ablation study on the number of receiver antennas used, to simulate ring designs with different numbers of antennas. Specifically, we randomly replace certain numbers (from 1 to 4) of channels on the recorded raw signal with Gaussian noise. We ran the ablation experiment using the leave-three-user-out data set. The results (Figure 9) show that using more antenna leads to higher testing accuracy, with an 28.5% increase from using one antenna to using five antennas.

User Impact. We also conducted an experiment to further investigate users' impact on their discrete gesture performance. Specifically, we experimented the impact of the user's finger size and age on the accuracy of classifying his/her discrete T2I microgestures. We trained 14 leave-one-user-out ViT models to test 14 users'
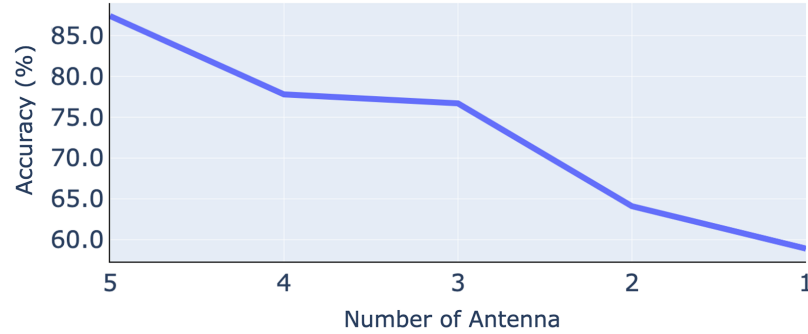
Fig. 9. Results of the ablation study on number of receiver antenna. The horizontal axes indicates number of antenna used. The vertical axes is the testing accuracy.

data respectively. For each leave-one-user-out ViT model, we trained it on 13 user's data for 64 epoch and tested the trained model on the left-out user's data. The average accuracy of these leave-one-user-out models is 81.1%. For each user, we calculated his/her thumb-to-index-finger length ratio, which is the ratio of his/her thumb width to the length of his/her index finger. Two Pearson product-moment correlation tests were conducted to investigate the user dependency of EFRing classifying discrete T2I microgestures. The results showed that there was no significant correlation between the user's thumb-to-index-finger length ratio and the gesture-classification performance of the leave-one-user-out model (r^2 = -0.50, p = 0.12), and no significant correlation between the user's age and the classification performance of the leave-one-user-out model (r^2 = 0.19, p = 0.58).

Thermal Drifting. We further conducted an experiment to investigate the potential impact of the heat dissipation of the hardware on the system performance. Specifically, we kept EFRing running for 10 hours and we asked 2 users who didn't participate in the previous studies to record 10 testing samples for each gesture in every 2 hours. The experiment was done in an in-door lab under a controlled temperature of $25°C$. In every 2 hours, we also measured the temperature from both the sensor chip and the ring surface. We then test the collected testing data on our leave-3-users-out model. Figure 10 shows that the temperature of the sensor chip (mean = $25.61°C$, SD = $1.22°C$) raised from $23.6°C$ to $27.0°C$, while the temperature of the ring (mean = $25.62°C$, SD = $0.10°C$) remain relatively stable throughout the 10-hours experiment period. The testing accuracy (mean = 85.35%, SD = 1.48%) in every 2 hours varied during the experiment period. The Pearson product-moment correlation tests showed that in every 2 hours, there was no significant correlation between the testing accuracy and the ring temperature (r^2 = -0.58, p = 0.23), and no significant correlation between the testing accuracy and the chip temperature (r^2 = 0.27, p = 0.60).

5.2.5 Experiments on Gesture Detection. In the real-world scenario, the EFRing sensing system should firstly detects whether a T2I microgesture is performed by the user, and then applies the classification model on the extracted gesture signal. For this purpose,

We randomly selected and re-invited twelve users from the previous study (7 males, 5 female, mane age = 25.75, SD = 1.82) for collecting additional "negative" data set of the label *NONE* with EFRing worn on their index fingers. To simulate a real-world usage scenario, we consider two "negative" situations: 1) the thumb is not in contact with the index finger; 2) the thumb is in contact with the index finger but staying still. For each situation,
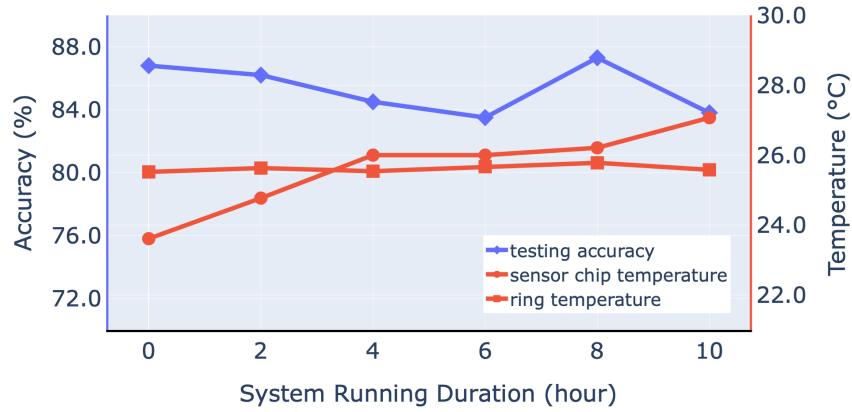
Fig. 10. Results of the thermal-drifting experiment. The horizontal axes indicates the device-running duration. The left vertical axes is the testing accuracy. The right vertical axes is the temperature. The purple line with diamond marker is the testing accuracy for every two hours. For the two red lines, the one with square marker denotes the measured ring temperature every two hours, while the one with circle marker indicates the measured sensor chip temperature in every two hours.

we record 5-minute signal data (57,000 frames of data under the sampling rate of 190Hz) from each person. The software and hardware setup was similar to the pilot experiment (Section 4.1). Upon the collected "negative" data, we simulate the real-world data stream by applying a 16ms (3-frames) sliding window across the data with the step size of 5ms (1 frame). For each window, we computed the channel-wise gradient (Section.5.2.3) as the data feature. The sliding window operation and feature extraction process yields 8043,753 data for training and 409,967 for testing, while each data is a vector with the dimension of $\mathbb{R}^{(10\times1)}$. We also randomly extract the the "positive" data samples (i.e. the signal that is recorded under a T2I microgesture) from our discrete-gesture data set (Section 5.2.1) in the same amount of the "negative" data. Noted that the "positive" data samples were chosen from the same twelve users where the "negative" data samples were recorded from.

With the "positive" and the "negative" data, we trained a binary SVM classifier. We employed the grid-searching mechanism to find the optimal set of hyper-parameters (RBF kernal, C = 1000, g = 0.1) for the SVM classifier. A five-fold cross-validation showed an average window-level classification accuracy of 96.5%. During the real-time detection, we applied a smoothing algorithm as Xu et al. [81] and Chen et al. [11] did. In particular, we treated adjacent sequences of continuous positive detection which lasted more than 1 seconds as a valid detection (i.e. 187 continuous positive windows under our setting). In practice, to reduce the noisy shifting, we tolerated if a long consecutive positive sequence was separated by few negative detection, and treated the whole sequence as a valid detection. We then used the segmented signal clips for future classification. To verify the real-time performance of our detection approach, we run a gesture detection experiment on our data set by simulating real-time data stream with sliding window algorithm across all data sample. Our experiment shows an 97.9% gesture-level detection accuracy with a recall of 97.2%.

6 CONTINUOUS T2I MOTION TRACKING

Besides the discrete microgestures, the T2I interaction space also supports the feasibility of continuous motion tracking [48]. Therefore, we investigate the feasibility of EFRing detecting and tracking continuous T2I subtle motions, such as dragging the thumb along the index finger as a virtual slider [72]. We proposed and evaluated
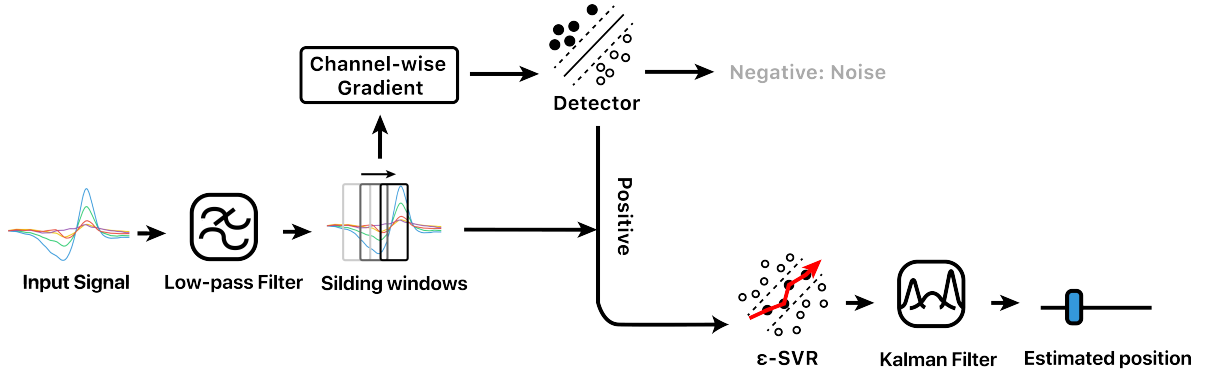
Fig. 11. The signal processing pipeline for our continuous 1D tracking algorithm. The system take raw data as input. A low-pass filter was applied to remove circuit noise. we applied sliding windows on the filtered data. We then calculate the channel-wise gradient on each sliding window before feeding to the gesture detector. If the gesture detector detect the user is performing gestures, we fit the raw window data to a ϵ-SVR for absolute position estimation. We further employed a customized *Kalman Filter* to smooth out the regression results.

an algorithm for robust and accurate continuous 1D tracking of the thumb's position on the index finger. We also evaluated the real-time performance and the user experience of EFRing tracking continous T2I motion through a 1D Fitts'-Law study.

6.1 Tracking Method

Equation 3 suggests that the output voltage from Rx(s) alters proportionally to the distance between the finger and Rx(s). As we slide the thumb between the distal and the proximal phalanx on index finger, the area between the thumb and the index finger, namely *purlicue circle*, will change as the thumb moving. The 5-channel signal from EFRing can reflect such space changes from the root of the index finger by measuring the change of Rx(s)'s output voltage. Such changes of Rx(s) output voltage could potentially be predicted with a regression-based algorithm, and be mapped to the thumb's position as it slides along the index finger.

Figure 11 shows the pipeline for tracking the thumb's position on the index finger using EFRing. Specifically, we build a machine-learning-based regressor, which takes a feature vector extracted from a 16ms (3-frames) sliding window of signal from EFRing as input and output an estimation of the current absolute position. In the implementation of real-time regression, we further employed a customized *Kalman Filter* [38] to smooth out the regression results. In addition, various hand sizes and finger lengths of different users could affect the ranges of the signal values while users moving the thumb along the index finger. For users who have a higher ratio of a user's thumb width to his/her index finger length (i.e. large thumb over short index finger), the moving range would be smaller, and the sensor signal range would also be smaller, and vice versa. To reduce the effect of this problem, we adopt a short "practice" section for the new user before he/she starts using EFRing. That is, we measure the extreme value of the new user by asking he/she to perform one complete sliding movement (i.e sliding from index finger tip to root and back to finger tip) for personal normalization/calibration. To this end, we can build a personalized profile by extracting the maximum and minimum sensor values for each channel, and normalize the new-incoming real-time signal using these range values. This "practice" process only needs to be executed once, as adopted by many commercial products (e.g., finger-print/face password).

6.2 Offline Experiments

To achieve the goal of continuous T2I motion tracking, we first conducted an offline experiments on comparing different types of regression models.

6.2.1 Data Collection. We recruited 12 participants (7 males, 5 females, mean age = 25.83, SD = 2.48) from a local university. All of them are right-handed, and without any prior experience on smart ring. The software and hardware apparatus are same as our aforementioned data-acquisition experiments. The experiment started after the participant filling a pre-questionnaire with demographic information and the measurement of his/her thumb-to-index-finger length ratio. The experiment facilitator first show a 600-pixel-long slider with its handle gradually moving in a constant speed (400 pixels/second) on the screen. The participants were asked to follow the moving handle by moving his/her thumb on the distal and middle phalanx of his/her index finger. They can practice until they felt confident on following the reference slider handle on the screen. Before start recording, the facilitator asked the participants to place his/her thumb on the tip of the distal phalanx as starting point. During recording, the references slider move from left to right and back to left in the constant speed of 400 pixels/second, and this is counted as one recording section. For every time step in each section, we recorded the signal from EFRing as well as the position of the reference slider handle as ground-truth values. Each participant completed two recroding sections where we treat one section as training data and the other one as testing data. The average time duration of one recording section across 12 participants is 542 frames (about 2.85 second).

6.2.2 Experimental Conditions. We performed an offline experiment on continuous T2I tracking by training and testing a series of regression models. To ensure the real-time performance of the proposed tracking method, we focused on comparing the light-weight regression models, including k-nearest neighbors (k-NN, k=5), Epsilon-Support Vector Regression (ϵ-SVR, RBF kernal, C=0.1, g=1), and a small multilayer perceptorn (MLP). Similar to the discrete gesture classification, we experimented these regression models under two training strategies: 1) experimenting the generic models trained and tested by the data from all the participants with a split of 2:1:1 for the training, the validation, and the testing data respectively; 2) experimenting the personalized models that are trained and tested from each participant only, also with the split of 2:1:1 (i.e. training : validation : testing). During the experiment, we simulate a real-world data stream by applying a 16ms sliding window across the signal data with the step size of 5ms.

6.2.3 Results. We calculated the mean squared error (MSE) of each regression model. To examine the smoothness of the simulated real-time tracking, we also compute the 2nd derivative of the regression result after applying our customized *Kalman Filter*. A smaller 2nd derivative value reveals a smoother regression results. Table 4(a) and (b) shows the result of our experiments on the generic models and the personalized models respectively. For the generic models, ϵ-SVR shows the smallest MSE, while MLP has the smoothest prediction. Although MLP shows the best smoothness, it relies on large data for a better convergence. In practice, it is reasonable to collect a small amount of data from a new user for calibration. Such calibration data can also be used for training a personalized model for a better performance. One-way ANOVA on the performance of the personalized model showed a significant difference among the three tested models in terms of MSE ($F(2,33)$ = 7.07, $p < 0.005$) and 2nd derivative ($F(2,33)$ = 25.80, $p < 0.0005$). The post-hoc pairwise comparisons show that the MSE of ϵ-SVR is significantly smaller than k-NN ($p < 0.05$) while there is no significant difference between ϵ-SVR and MLP. The 2nd derivative value of k-NN is significantly larger than ϵ-SVR ($p < 0.005$) and MLP ($p < 0.005$). As a result, we use the personalized ϵ-SVR for our user study on real-time T2I motion tracking.

6.3 Usability Study

We also conducted a user study to investigate the real-time performance and usability of 1D T2I motion tracking with EFRing.

Table 4. Results for the regression experiment. This table summarizes the mean squared error (MSE) as well as the second derivative of the regression result after applying *Kalman Filter* for three different models. In table (a), all the models were trained and tested on the data from all users. In table (b), all the models were separately trained and tested on individual user data, and the results are averaged across 12 personalized models.

(a) Regression results on the generic models.

Models	k-NN	ϵ-SVR	MLP
MSE↓	3.8%	3.5%	4.8%
2nd Derivative↓	1.24×10^{-3}	1.77×10^{-4}	9.39×10^{-5}

(b) Regression results on the personalized models.

Models	k-NN	ϵ-SVR	MLP
MSE↓	3.5%	2.3%	4.5%
2nd Derivative↓	8.26×10^{-4}	2.74×10^{-4}	3.13×10^{-4}

(1/60) Start

GESTURING

Fig. 12. The interface of our fitts' law study. Users are required to move the slider handle into the target area. Once success, the target ares will shows in red.

6.3.1 Participants and Apparatus. All the twelve participants in the data-acquisition study (Section 6.2.1) were invited back to evaluate the system. The software for this experiment was developed with Qt5.6 framework in C++, where we implemented a ϵ-SVR model with LIBSVM library [9]. A customized *Kalman Filter* was implemented using the Eigen library [27]. A real-time gesture detector (Section 5.2.5) was also implemented in this study. All the participants are required to wear EFRing on the proximal phalanx of their right-hand index fingers during the experiment.

6.3.2 Task and Metrics. Fitts' Law [16] has been widely used for evaluating a input interface by modeling human motor behaviour [4, 53]. In Fitts' Law, the task's index of difficulty (*ID*, in bits) is defined as:

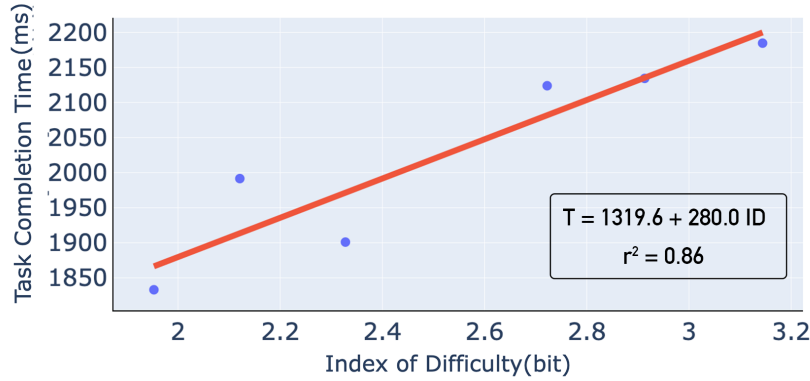$$ID = \log_2(\frac{A}{W} + 1) \tag{8}$$

Fig. 13. Fitts' Law Regression Result. The x axes is the index of difficulty (ID, in bits), while the y axes is the task completion time (T, in ms). The purple points are results for different $A \times W$ conditions from our study, while the red line is the linear regression result to fit them.

where A is the distance from the starting point to the target, W is the width of the target region. Fitts' Law indicates the correlation between the task-completion time T and the task's index of difficulty ID through a linear equation as below:

$$T = a + bID \tag{9}$$

where a and b are the regression coefficient. The performance of task can be evaluated by solving the linear-regression problem and examining the correlation of determination (r^2) for the goodness of fit.

We designed a 1D target-selection task, moving the handle to select the rectangular target on the sliding bar, for our study. We included 6 $A \times W$ conditions, with 2 levels of A (201 and 392 pixels) and 3 level of W (50, 60, and 70 pixels), yielding 6 ID ranging from 1.95 bits to 3.13 bits. A was measured from the left end of the sliding bar to the center of the target. W was the width of the rectangular target. Each combination of $A \times W$ repeated for 10 times.

During the study, the participants wearing EFRing can control the slider handle by moving his/her thumb on the index finger. Before each trial, the experiment facilitator first asked participants to place his/her thumb on the tip of the index finger as the starting point for calibration. At the beginning of each trial, as shown in (Fig. 13), there was a target in a specific width A appeared at a specific position D on the sliding bar. As the trial starts, the participants were instructed to move the slider handle from the starting point to the target as fast and as accurate as possible. Upon successfully selecting the target region (i.e. the slider handle hits the target), participants were told to lift his/her thumb from the index finger for conformation. We measure the time duration between the task-starting timestamp and the participant-conformation timestamp, as the task-completion time.

After the experiment, the participant completed a user-experience questionnaire [52] on their experience (1 - strongly disagree, 7 - strongly agree) of the system.

6.3.3 Results. We drew a quantitative analysis on our Fitts' Law study as well as a qualitative analysis on the user-experience questionnaire.

Fitts' Law Analysis. We compute the average time consumption for every $A \times W$ combination across different users. The average task-completion time per trial is 2028.05ms (SD = 141.83). Figure 13 shows the index of performance of 1D target selection using EFRing, by plotting the result of linear regression between between

task completion time and index of difficulty (ID). The regression statistic shows the r^2 is 0.86 (p < 0.05) which is closed to the results of previous research on 1D Fitts' Law [4], inferring that users using EFRing could potentially achieve similar performance on 1D target selection as they did on the touch screen with their fingers. We also performed the Pearson product-moment correlation on the participants' thumb-to-index-finger length ratio and their task-completion time, to investigate the user dependency of our approach. The results showed that The Fitts'-law task-completion time was strongly negatively correlated to this ratio (r^2 = -0.60, p < 0.05), indicating the users with shorter index fingers tended to finish the selection task faster. This follows the assumption of Fitts' Law that shorter distance yields shorter reaching time.

Usability and User Engagement. The questionnaire results (Figure 14) show that participants rated the system is *easy to learn* (6.33/7, SD=0.49) and *easy to use* (6.25/7, SD=0.62). The engagement of using EFRing for target selection achieved an average score of 6.58/7 (SD = 0.67), and the intuitiveness of the interaction is rated averagely 6.58/7 (SD = 0.51). Two participants reported that it was *"easy to get familiar with the system".* Two other participants commented that *"the experience is super impressive.",* and one participants explicitly mentioned that *"it is amazing to control the slider without touch the screen"* and she *"can't wait to recommend it to her friends".* Another participant mentioned that he *"feel very confident to use the system."*

7 EFRING APPLICATIONS

As shown in our experiments, EFRing offers a reliable and real-time input method for both discrete microgesture classification and continuous motion tracking of T2I interaction. In this section, we present three potential applications: Google Cardboard VR interaction, 2D sketching on the touch surface, and device control. Noted that the presented applications were distilled from our experiment participants' comments on how EFRing could be used in the real-life scenario.

7.1 Google Cardboard VR Interaction

While Google Cardboard VR offers a low-cost and immersive VR experience to the masses, its interactivity is still limited to either finger tapping on the conductive tape or head orientation most of the time [11]. This limitation may affect the user experience in many VR applications such as gaming, viewing 3D objects, and video playback control.

In our proof-of-concept mobile applications, we connected EFRing to an Android smartphone wirelessly through Bluetooth. Therefore, EFRing can serve as an input device for various VR scenarios of Google Cardboard. For example, directional discrete gesture such as *Swipe Up*, *Swipe Down*, *Swipe Left*, and *Swipe Right* could be used for 3D world navigation in mobile VR (Fig. 1e). Combined with *Tap* or *Double Tap*, EFRing also could facilitate menu navigation in mobile VR (Fig. 1f). Continuous controls such as rotating 3D objects and rolling video playback progress bars can also be supported by the continuous sliding gesture with EFRing (Fig. 1g).

While we mainly focus on the thumb-to-index-finger microgesture interaction that is not directly on the ring device (i.e. off-ring interaction), existing smart-ring research has shown the feasibility of on-ring interaction with various sensors (e.g., capacitive touch sensors [5, 70], and buttons [2]). Combining the off-ring interaction supported by EFRing and the existing on-ring interaction could enrich the interaction space of smart-ring devices.

7.2 2D Sketching on the Touch Screen

EFRing could serve as an alternative input method to the GUI-based control on the touch screen. Fig 1g shows an application example on using EFRing for parameter adjustment in a 2D-sketching task on the touch screen. EFRing's continuous sliding gesture could be used to control the sketching parameters, such as the stroke weight and the color. In such an application, artists can adjust these parameters while drawing, instead of lifting their fingers from the screen and touching the GUI elements to toggle the parameter-adjustment mode. Besides, the

System Usability Scale Results

I think that I would like to use this system frequently

I found the system unnecessarily complex

I think the system was easy to use

I think that I would need the support of a technical person to be able to use this system

I found the various functions in this system were well integrated

I thought there was too much inconsistency in this system

I would imagine that most people would learn to use this system very quickly

I found the system very cumbersome to use

I felt very confident using the system

I needed to learn a lot of things before I could get going with this system

User Experience Questionnaires Results

This input method is intuitive

This input method is fun

I can precisely control the cursor using the system

I can easily control the cursor using this system

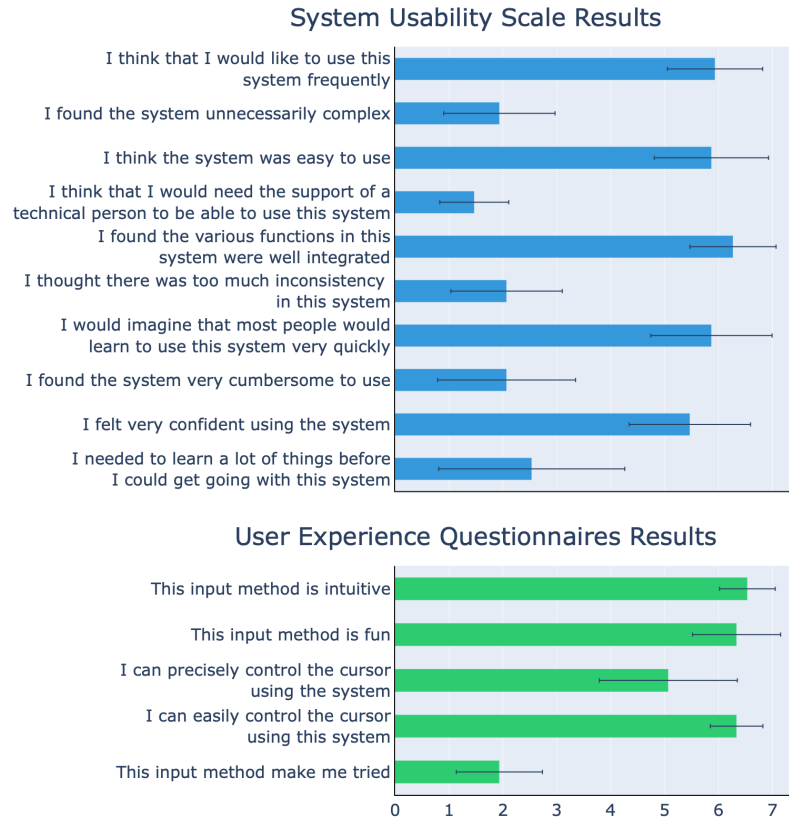This input method make me tried

0 1 2 3 4 5 6 7

Fig. 14. The questionnaire results of our usability study, including standard system usability scale (SUS) and customized questionnaires on the user experiences of our system.

discrete T2I gestures can also serve as shortcut commands for mode switching. For example, the *Circle Clockwise* gesture can represent as the undo command, while the *Circle Counterclockwise* gesture is the redo command. *Swipe Up* and *Swipe Down* can be used for switching between different sketching tools such as brush and eraser.

Besides the 2D sketching application, the similar control concept can be adopted to the other GUI-based controls. For example, in a text-editing application, user can highlight or change the style of the text using T2I microgesture after he/she drags and selects the text. Continuous motion tracking could be used for tuning the font size or changing the text color.

7.3 Device Control

With EFRing, users can interact with other smart devices, such as mobile phones, in a private and unobtrusive way. For instance, during a meeting or in a library, the discrete T2I gestures recognized by EFRing can be used to hang up the incoming phone or handle the notifications quietly. In the context of IoT, users can use the T2I continous motions supported by EFRing to control the smart-home devices. For instance, adopting the device-selection method using finger pointing [13], users can adjust the room brightness/temperature or the music volume by sliding their thumbs on their index fingers after selecting a particular home device.

8 LIMITATIONS AND FUTURE WORK

In this paper, we focus on the EF-sensing technique and the signal-processing algorithms for supporting T2I microgesture interaction with only one index-finger ring. Although we show the feasibility of EFRing for detecting a various of subtle thumb movement patterns including discrete gestures and continuous motion, there are still remaining some limitations and potentials for future improvements.

Real-world Experiments. The presented EFRing experiments mainly focused on the context of users sitting still indoor without carrying any item in their hands. Existing research shows that the user experience of the smart ring could be affected by the body movement (e.g., walking and running) and the hand cumber (i.e., carrying heavy items) [64]. Therefore, it is predictable that the EFRing signals induced by the T2I microgestures could be different in these different contexts. As an important future work, it is worth to investigate gesture classification and tracking with EFRing in different real-world contexts: walking in different speeds, and carrying items in different weights.

Non-bare-hand Usage Scenarios. We mainly focused on the bare-hand situation of EFRing wearing. While the external interference of the EF-sensing signal is mainly from the conductive objects, it is unknown how the hand-worn accessories may affect the signal quality for gesture classification and tracking. For instance, when a user is wearing the glove on the EFRing-wearing hand, the glove will cover the device and the fingers, and the coverage may affect the finger movement and the signal features. In another scenario of hand-worn accessory, the metal rings on the other fingers may interfere the EF-sensing signal around the index finger.

Impact of Environment. Our experiments were done in a in-door lab under a controlled temperature of $25°C$. It could be possible that the electric field would vary under different temperatures due to the change of ε_0 (Equation. 2). According to Equation 3, the influence of ε_0 will be theoretically neglected while calculating the final output voltage V_o^k. However, more experiments need to be done in the future to further verify the sensing capability of EFRing in different environmental conditions. In addition, it could also be arguable that the electric field would be affected by any nearby conductive objects. In practice, a passive conductive object, such as metallic rings worn on the fingers, would generate a constant interference on the EF signal. Noted that we preprocessed the signal by calculating the temporal-wise gradient and the channel-wise gradient for detecting both discrete and continuous T2I gestures. Such local gradient-based features would not be affected by any global constant influence. Therefore, we could argue that our solution is robust for any static nearby conductive objects. We plan to conduct more experiments on it to further verify our assumption.

Comparison with GUI. With the T2I microgestures enabled by EFRing, it would be necessary to explore the mappings between the EFRing-enabled gestures and the real-world applications, and investigate the effectiveness of these gestures on different interaction tasks. Previous research showed that gesture-based interaction outperformed graphical-user-interface-based (GUI-based) interaction by yielding shorter task-completion time and lower workload [82]. We hypothesize that EFRing may yield a similar better performance over GUI-based interaction in the applications such as VR navigation, touch-screen 2D sketching, etc. We plan to conduct a thorough usability experiment in the near future.

Mode Integration. Our experiments showed the feasibility of discrete gesture classification and continuous motion tracking using EFRing, but the current implementation and experiments treated these two types of T2I interaction as two separated modes for different applications. In the real-world scenario, it is possible that the user may switch between these two modes to fulfill different interaction tasks. Therefore, there is a need on designing the mode-switching mechanism between discrete gesture classification and continuous motion tracking. In addition, there would be a potential need for the activation technique (e.g., on-ring tapping) for starting the classification/tracking pipeline, to avoid false-positive detection.

Work as An Untethered Device. The current prototype of EFRing microgesture classification and tracking was run on a desktop PC as a proof of concept. The performance could be negatively affected while directly running the machine-learning models on the mobile devices (e.g., smart phone/watch/glasses) due to the model complexity and the computational constraint on the device. While this issue could be solved with the future hardware advancing in the smart devices, one potential solution that can be feasible in the near future is leveraging the advantage of modern high-speed mobile networks (e.g., 5G cellular network) [55, 86]. Guo Tian [28] suggested that the modern deep-learning-based mobile applications could benefit from the hybrid approach of combining the on-device and the cloud-based classification. More specifically for EFRing, we could run the light-weight gesture-detection and the continuous motion-tracking process in real-time locally on the smart devices. For discrete gesture classification, the EF-sensing signal could be simultaneously sent to the cloud server through the cellular network (e.g., 5G) in low latency.

While our current experiments were done with a tethered device as a proof of concept, it can be argued that the quality of the EF signal (e.g. SNR) may decrease if the device is not connected to the earth ground (i.e. as a mobile device), namely in untethered mode [54]. To reduce the signal noise in our current prototype, we increased the grounded area of the PCB by covering a grounded copper pour to reduce high-frequency signal noise. In the software part, we implemented a low-pass noise filter before extracting signal features to suppress the high-frequency noise. To test the performance of EFRing in the untethered mode, we conducted an informal preliminary test, where we recruited 2 male users (25 and 26 years old) to record 40 testing samples for each discrete microgestures with EFRing connected to an unplugged laptop. The testing experiments on the trained ViT model resulted in an average accuracy of 83.5%, a 1.7% drop compared to our aforementioned experiments with the testing data collected from the tethered device. As suggested by Matthies et al. [54], the signal quality in the untethered standalone device could be improved by either enhancing the capacitive coupling between the human body and the sensing antennas or creating a small electric field around the body and the electrode. We will experiment with these approaches for EFRing in-depth in the future.

Robustness of Gesture Detection. In this paper, we mainly focus on sensing both discrete and continuous gestures with EFRing. However, the gesture detection, as a system activation process, is another important component for a real-time always-on sensing system in real-world usage. In section 5.2.5, we conducted a proof-of-concept experiments on the gesture-detection algorithm. The algorithm was also deployed in real-time for our usability study (Section 6.3). Both studies showed that our current detection algorithm could perform satisfactorily in the laboratory environment. However, as a real-world usage scenario could be more complex compared with the laboratory setting, our detection system could be improved in two aspects. On one hand, as we only considered the *NONE* case when the user sat still, more usage scenarios (e.g. walking, running, carrying heavy items, etc.) needed to be included as the future work. On the other hand, we collected our "positive" and "negative" samples in two separate sections, which may introduce environmental bias in our data set. We will further verify our detection algorithm on a larger and more balanced data set.

Beyond 1D Continuous Tracking. During our experiments, we showed that EFRing is capable for detecting 1D continuous movements of T2I motion, in the direction parallel to the distal phalanx of the index finger. The width of the phalanx is usually shorter than its length, leading to lower resolution of the thumb motion on the direction perpendicular to the distal phalanx of the index finger. This places extra challenges on tracking the thumb motion on this direction and 2D T2I motion. Tracking 3D thumb motion may involve mid-air gestures which could evoke unique signals due to the capacitive nature of EF sensing. In the future work, we will experiment new design of antenna layout and EF signals to explore 2D/3D thumb motion tracking.

9 CONCLUSION

In this paper, we present EFRing, an index-finger-worn ring-form device for detecting T2I microgestures through EF sensing. Our experiments show that the signals recorded by EFRing could be used for both discrete microgesture classification and continuous motion tracking. The ViT-based classification showed an average within-user accuracy of 89.2% and an average cross-user accuracy of 85.2%, for nine T2I microgestures. For the continuous T2I continuous motion tracking, our ϵ-SVR-based method can achieve the mean-square error of 3.5% for the generic model and 2.3% for the personalized model. The 1D-fitts'-law target-selection study shows that the proposed tracking method with EFRing is intuitive and accurate. With EFRing, we demonstrate the feasibility of T2I microgesture interaction through only one index-finger-wearable device, and we hope to enrich the interaction paradigm for micro-space user interfaces.

ACKNOWLEDGMENTS

REFERENCES

[1] Kazuyuki Arimatsu and Hideki Mori. 2020. Evaluation of Machine Learning Techniques for Hand Pose Estimation on Handheld Device with Proximity Sensor. *Conference on Human Factors in Computing Systems - Proceedings* (2020), 1–13. https://doi.org/10.1145/3313831.3376712

[2] Sandra Bardot, Surya Rawat, Duy Thai Nguyen, Sawyer Rempel, Huizhe Zheng, Bradley Rey, Jun Li, Kevin Fan, Da Yuan Huang, Wei Li, and Pourang Irani. 2021. ARO: Exploring the Design of Smart-Ring Interactions for Encumbered Hands. *Proceedings of MobileHCI 2021 - ACM International Conference on Mobile Human-Computer Interaction: Mobile Apart, MobileTogether* (2021). https://doi.org/10.1145/3447526.3472037

[3] Hymalai Bello, Bo Zhou, Sungho Suh, and Paul Lukowicz. 2020. MoCapaci: Posture and gesture detection in loose garments using textile cables as capacitive antennas. *Proceedings - International Symposium on Wearable Computers, ISWC* (2020), 78–83. https://doi.org/10.1145/3460421.3480418

[4] Xiaojun Bi, Yang Li, and Shumin Zhai. 2013. FFitts Law: Modeling finger Touch with Fitts' Law. *Conference on Human Factors in Computing Systems - Proceedings* (2013).

[5] Roger Boldu, Alexandru Dancu, Denys J.C. Matthies, Pablo Gallego Cascón, Shanaka Ransir, and Suranga Nanayakkara. 2018. Thumb-in-motion: Evaluating thumb-to-ring microgestures for athletic activity. *SUI 2018 - Proceedings of the Symposium on Spatial User Interaction* (2018), 150–157. https://doi.org/10.1145/3267782.3267796

[6] T. Caliński and J Harabasz. 1974. A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods* 3, 1 (1974), 1–27.

[7] Liwei Chan, Yi Ling Chen, Chi Hao Hsieh, Rong Hao Liang, and Bing Yu Chen. 2015. Cyclopsring: Enabling whole-hand and context-aware interactions through a fisheye ring. *UIST 2015 - Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology* 1 (2015), 549–556. https://doi.org/10.1145/2807442.2807450

[8] Liwei Chan, Rong-Hao Liang, Ming-Chang Tsai, Kai-Yin Cheng, Chao-Huai Su, Mike Y. Chen, Wen-Huang Cheng, and Bing-Yu Chen. 2013. FingerPad: Private and Subtle Interaction Using Fingertips. (2013), 255–260. https://doi.org/10.1145/2501988.2502016

[9] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2 (2011), 27:1–27:27. Issue 3. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[10] Ke-Yu Chen, Shwetak N. Patel, and Sean Keller. 2016. Finexus: Tracking Precise Motions of Multiple Fingertips Using Magnetic Sensing. *Conference on Human Factors in Computing Systems - Proceedings* (2016), 1504–1514. https://doi.org/10.1145/2858036.2858125

[11] Taizhou Chen, Lantian Xu, Xianshan Xu, and Zhu Kening. 2021. GestOnHMD : Enabling Gesture-based Interaction on Low-cost VR Head-Mounted Display. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2597–2607.

[12] Gabe Cohn, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. Humantenna: Using the Body as an Antenna for Real-Time. *Conference on Human Factors in Computing Systems - Proceedings* (2012), 1901–1910. https://doi.org/10.1145/2207676.2208330

[13] Rajkumar Darbar. 2019. RingIoT : A Smart Ring Controlling Things in Physical Spaces. May (2019), 2–9.

[14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

[15] Christoph Endres, Tim Schwartz, and Christian Müller. 2011. "Geremin": 2D microgestures for drivers based on electric field sensing. *International Conference on Intelligent User Interfaces, Proceedings IUI* June 2014 (2011), 327–330. https://doi.org/10.1145/1943403.1943457

[16] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.

[17] Masaaki Fukumoto and Yasuhito Suenaga. 1994. "FingeRing": A full-time wearable interface. *Conference on Human Factors in Computing Systems - Proceedings* 1994-April, May (1994), 81–82. https://doi.org/10.1145/259963.260056

[18] Oliver Glauser, Daniele Panozzo, Otmar Hilliges, and Olga Sorkine-Hornung. 2019. Deformation capture via self-sensing capacitive arrays. *ACM Trans. Graph* 38, 2 (2019), 1–16.

[19] Oliver Glauser, Shihao Wu, Daniele Panozzo, Otmar Hilliges, and Olga Sorkine-Hornung. 2019. Interactive Hand Pose Estimation using a Stretch-Sensing Soft Glove. *SACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)* 38, 4 (2019), 15. https://doi.org/10.1145/3306346.3322957

[20] Mathieu Le Goc, Stuart Taylor, Shahram Izadi, and Cem Keskin. 2014. A lowcost transparent electric field sensor for 3D interaction on mobile devices. *Conference on Human Factors in Computing Systems - Proceedings* April (2014), 3167–3170. https://doi.org/10.1145/2556288.2557331

[21] Jun Gong, Yang Zhang, Xia Zhou, and Xing Dong Yang. 2017. Pyro: Thumb-tip gesture recognition using pyroelectric infrared sensing. *UIST 2017 - Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (2017), 553–563. https://doi.org/10.1145/3126594.3126615

[22] Tobias Grosse-Puppendahl, Yannick Berghoefer, Andreas Braun, Raphael Wimmer, and Arjan Kuijper. 2013. OpenCapSense: A rapid prototyping toolkit for pervasive interaction using capacitive sensing. *2013 IEEE International Conference on Pervasive Computing and Communications, PerCom 2013* March (2013), 152–159. https://doi.org/10.1109/PerCom.2013.6526726

[23] Tobias Grosse-Puppendahl and Andreas Braun. 2012. Honeyfish - a high resolution gesture recognition system based on capacitive proximity sensing. *Embedded World Conference 2012* January (2012), 10pp.

[24] Tobias Grosse-Puppendahl, Christian Holz, Gabe Cohn, Raphael Wimmer, Oskar Bechtold, Steve Hodges, Matthew S. Reynolds, and Joshua R. Smith. 2017. Finding common ground: A survey of capacitive sensing in human-computer interaction. *Conference on Human Factors in Computing Systems - Proceedings* 2017-May (2017), 3293–3316. https://doi.org/10.1145/3025453.3025808

[25] Changzhan Gu and Jaime Lien. 2017. A Two-Tone Radar Sensor for Concurrent Detection of Absolute Distance and Relative Movement for Gesture Sensing. *IEEE Sensors Letters* 1, 3 (2017), 1–4. https://doi.org/10.1109/lsens.2017.2696520

[26] Yizheng Gu, Chun Yu, Zhipeng Li, Weiqi Li, Shuchang Xu, Xiaoying Wei, and Yuanchun Shi. 2019. Accurate and low-latency sensing of touch contact on any surface with finger-worn IMU sensor. *UIST 2019 - Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (2019), 1059–1070. https://doi.org/10.1145/3332165.3347947

[27] Gaël Guennebaud, Benoît Jacob, et al. 2010. Eigen v3. http://eigen.tuxfamily.org.

[28] Tian Guo. 2018. Cloud-based or on-device: An empirical study of mobile deep inference. In *2018 IEEE International Conference on Cloud Engineering (IC2E)*. IEEE, 184–190.

[29] Shangchen Han, Beibei Liu, Randi Cabezas, Christopher D. Twigg, Peizhao Zhang, Jeff Petkau, Tsz Ho Yu, Chun Jung Tai, Muzaffer Akbay, Zheng Wang, Asaf Nitzan, Gang Dong, Yuting Ye, Lingling Tao, Chengde Wan, and Robert Wang. 2020. MEgATrack: Monochrome Egocentric Articulated Hand-Tracking for Virtual Reality. *ACM Transactions on Graphics* 39, 4 (2020). https://doi.org/10.1145/3386569.3392452

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

[31] Mikko Heino, Dani Korpi, Timo Huusari, Emilio Antonio-Rodriguez, Sathya Venkatasubramanian, Taneli Riihonen, Lauri Anttila, Clemens Icheln, Katsuyuki Haneda, Risto Wichman, et al. 2015. Recent advances in antenna design and interference cancellation algorithms for in-band full duplex relays. *IEEE Communications magazine* 53, 5 (2015), 91–101.

[32] Anuradha Herath, Bradley Rey, Sandra Bardot, Sawyer Rempel, Lucas Audette, Huizhe Zheng, Jun Li, Kevin Fan, Da-Yuan Huang, Wei Li, et al. 2022. Expanding Touch Interaction Capabilities for Smart-rings: An Exploration of Continual Slide and Microroll Gestures. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.

[33] Fang Hu, Peng He, Songlin Xu, Yin Li, and Cheng Zhang. 2020. FingerTrak: Continuous 3D Hand Pose Tracking by Deep Learning Hand Silhouettes Captured by Miniature Thermal Cameras onWrist. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020). https://doi.org/10.1145/3397306

[34] Da Yuan Huang, Liwei Chan, Shuo Yang, Fan Wang, Rong Hao Liang, De Nian Yang, Yi Ping Hung, and Bing Yu Chen. 2016. Digitspace: Designing Thumb-to-fingers touch interfaces for one-handed and eyes-free interactions. *Conference on Human Factors in Computing Systems - Proceedings* (2016), 1526–1537. https://doi.org/10.1145/2858036.2858483

[35] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

[36] Seungwoo Je, Minkyeong Lee, Yoonji Kim, Liwei Chan, Xing-Dong Yang, and Andrea Bianchi. 2018. PokeRing: Notifications by Poking Around the Finger. *Conference on Human Factors in Computing Systems - Proceedings* (2018), 1–10. https://doi.org/10.1145/3173574.3174116

[37] Seungwoo Je, Brendan Rooney, Liwei Chan, and Andrea Bianchi. 2017. tactoRing: A skin-drag discrete display. *Conference on Human Factors in Computing Systems - Proceedings* 2017-May (2017), 3106–3114. https://doi.org/10.1145/3025453.3025703

[38] Rudolph Emil Kalman. 1960. A new approach to linear filtering and prediction problems. (1960).

[39] Wolf Kienzle and Ken Hinckley. 2014. LightRing: Always-Available 2D Input on Any Surface. *UIST 2014 - Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (2014), 157–160. https://doi.org/10.1145/2642918.2647376

[40] Wolf Kienzle, Eric Whitmire, Chris Rittaler, and Hrovje Benko. 2021. Electroring: Subtle pinch and touch detection with a ring. *Conference on Human Factors in Computing Systems - Proceedings* (2021). https://doi.org/10.1145/3411764.3445094

[41] David Kim, Otmar Hilliges, Shahram Izadi, Alex Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: Freehand 3D interactions anywhere using a wrist-worn gloveless sensor. *UIST'12 - Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (2012), 167–176.

[42] Jonghwa Kim, Stephan Mastnik, and Elisabeth André. 2008. EMG-based hand gesture recognition for realtime biosignal interfacing. *International Conference on Intelligent User Interfaces, Proceedings IUI* (2008), 30–39. https://doi.org/10.1145/1378773.1378777

[43] Diederik P. Kingma and Jimmy Lei Ba. 2015. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (2015), 1–15. arXiv:1412.6980

[44] George Frederick Kunz. 2012. *Rings for the Finger.* Courier Corporation.

[45] Gierad Laput and Chris Harrison. 2019. SurfaceSight: A New Spin on Touch, User, and Object Sensing for IoT Experiences. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.*

[46] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers. *UIST 2016 - Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (2016), 321–333. https://doi.org/10.1145/2984511.2984582

[47] Chen Liang, Chun Yu, Yue Qin, Yuntao Wang, and Yuanchun Shi. 2021. DualRing: Enabling Subtle and Expressive Hand Interaction with Dual IMU Rings. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021). https://doi.org/10.1145/3478114

[48] Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: Ubiquitous Gesture Sensing with Millimeter Wave Radar. *ACM Transactions on Graphics* 35, 4 (2016), 1–19. https://doi.org/10.1145/2897824.2925953

[49] Hyunchul Lim, Jungmin Chung, Changhoon Oh, SoHyun Park, Joonhwan Lee, and Bongwon Suh. 2018. Touch+ Finger: Extending Touch-based User Interface Capabilities with" Idle" Finger Gestures in the Air. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology.* 335–346.

[50] Guanhong Liu, Yizheng Gu, Yiwen Yin, Chun Yu, Yuntao Wang, Haipeng Mi, and Yuanchun Shi. 2020. Keep the Phone in Your Pocket: Enabling Smartphone Operation with an IMU Ring for Visually Impaired People. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020). https://doi.org/10.1145/3397308

[51] Yilin Liu, Shijia Zhang, and Mahanth Gowda. 2021. NeuroPose: 3D hand pose tracking using EMG wearables. *The Web Conference 2021 - Proceedings of the World Wide Web Conference, WWW 2021* (2021), 1471–1482. https://doi.org/10.1145/3442381.3449890

[52] Arnold M Lund. 2001. Measuring usability with the USE questionnaire. *Usability interface* 8, 2 (2001), 3–6.

[53] I Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction* 7, 1 (1992), 91–139.

[54] Denys JC Matthies, Chamod Weerasinghe, Bodo Urban, and Suranga Nanayakkara. 2021. CapGlasses: Untethered Capacitive Sensing with Smart Glasses. In *Augmented Humans Conference 2021.* 121–130.

[55] Miranda McClellan, Cristina Cervelló-Pastor, and Sebastià Sallent. 2020. Deep learning at the mobile edge: Opportunities for 5G networks. *Applied Sciences* 10, 14 (2020), 4735.

[56] MicroChip. 2022. *ATUSB-GESTIC-PCB.* Retrieved Aug 10, 2022 from https://www.microchip.com/en-us/development-tool/EV91M41A

[57] MicroChip. 2022. *MGC3130 Datasheet.* Retrieved Aug 10, 2022 from https://www.microchip.com/en-us/product/MGC3130

[58] Franziska Mueller, Micah Davis, Florian Bernard, Oleksandr Sotnychenko, Mickeal Verschoor, Miguel A. Otaduy, Dan Casas, and Christian Theobalt. 2019. Real-time pose and shape reconstruction of two interacting hands with a single depth camera. *ACM Transactions on Graphics* 38, 4 (2019). https://doi.org/10.1145/3306346.3322958 arXiv:2106.08059

[59] Adiyan Mujibiya. 2013. Mirage : Exploring Interaction Modalities Using Off - Body Static Electric Field Sensing. *UIST 2013 - Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (2013), 211–220.

[60] Arshad Nasser, Taizhou Chen, Can Liu, Kening Zhu, and PVM Rao. 2020. FingerTalkie: Designing a Low-Cost Finger-Worn Device for Interactive Audio Labeling of Tactile Diagrams. In *Human-Computer Interaction. Multimodal and Natural Interaction*, Masaaki Kurosu

(Ed.). Springer International Publishing, Cham, 475–496.

[61] Viet Nguyen, Siddharth Rupavatharam, Luyang Liu, Richard Howard, and Marco Gruteser. 2019. HandSense: Capacitive coupling-based dynamic, micro finger gesture recognition. *SenSys 2019 - Proceedings of the 17th Conference on Embedded Networked Sensor Systems* (2019), 285–297. https://doi.org/10.1145/3356250.3360040

[62] J. Rekimoto. 2001. GestureWrist and GesturePad: Unobtrusive wearable interaction devices. *International Symposium on Wearable Computers, Digest of Papers* (2001), 21–27. https://doi.org/10.1109/iswc.2001.962092

[63] Thijs Roumen, Simon T. Perrault, and Shengdong Zhao. 2015. NotiRing: A Comparative Study of Notification Channels for Wearable Interactive Rings. *Conference on Human Factors in Computing Systems - Proceedings* (2015), 2497–2500. https://doi.org/10.1145/2702123.2702350

[64] Shardul Sapkota, Ashwin Ram, and Shengdong Zhao. 2021. Ubiquitous Interactions for Heads-Up Computing: Understanding Users' Preferences for Subtle Interaction Techniques in Everyday Settings. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse amp; Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 36, 15 pages. https://doi.org/10.1145/3447526.3472035

[65] Yilei Shi, Haimo Zhang, Kaixing Zhao, Jiashuo Cao, Mengmeng Sun, and Surang Nanayakkara. 2020. Ready, Steady, Touch! — Sensing Physical Contact with a Finger-Mounted IMU. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4 (2020), 1—-25.

[66] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[67] Joshua Smith, Tom White, Christopher Dodge, Joseph Paradiso, Neil Gershenfeld, and David Allport. 1998. Electric field sensing for graphical interfaces. *IEEE Computer Graphics and Applications* 18, 3 (1998), 54–59. https://doi.org/10.1109/38.674972

[68] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2016-December (2016), 2818–2826. https://doi.org/10.1109/CVPR.2016.308 arXiv:1512.00567

[69] Jonathan Tompson, Murphy Stein, Yann Lecun, and Ken Perlin. 2014. Real-time continuous pose recovery of human hands using convolutional networks. *ACM Transactions on Graphics* 33, 5 (2014). https://doi.org/10.1145/2629500

[70] Hsin-Ruey Tsai, Min-Chieh Hsiu, Jui-Chun Hsiao, Lee-Ting Huang, Mike Chen, and Yi-Ping Hung. 2016. TouchRing: Subtle and Always-Available Input Using a Multi-touch Ring. In *The ACM International Conference on Mobile Human-Computer Interaction*. 891–898. https://doi.org/10.1145/2957265.2961860

[71] Hsin Ruey Tsai, Lee Ting Huang, Cheng Yuan Wu, and Yi Ping Hung. 2016. ThumbRing: Private interactions using one-handed thumb motion input on finger segments. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, MobileHCI 2016* (2016), 791–798. https://doi.org/10.1145/2957265.2961859

[72] Hsin-Ruey Tsai, Te-Yen Wu, Da-Yuan Huang, Min-Chieh Hsiu, Jui-Chun Hsiao, Yi-Ping Hung, Mike Y. Chen, and Bing-Yu Chen. 2017. SegTouch: Enhancing Touch Input While Providing Touch Gestures on Screens Using Thumb-To-Index-Finger Gestures. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (2017), 2164–2171. https://doi.org/10.1145/3027063.3053109

[73] Andre J Van Schyndel and Diane J Clayton. 2005. Electric field proximity detector for floating and grounded targets. US Patent 6,859,141.

[74] Radu-Daniel Vatavu and Laura-Bianca Bilius. 2021. *GestuRING: A Web-Based Tool for Designing Gesture Input with Rings, Ring-Like, and Ring-Ready Devices.* Association for Computing Machinery, New York, NY, USA, 710–723. https://doi.org/10.1145/3472749.3474780

[75] Robert Y. Wang and Jovan Popović. 2009. Real-time hand-tracking with a color glove. *ACM Transactions on Graphics* 28, 3 (2009), 1–8. https://doi.org/10.1145/1531326.1531369

[76] Saiwen Wang, Jie Song, Jamie Lien, Ivan Poupyrev, and Otmar Hilliges. 2016. Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. *UIST 2016 - Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (2016), 851–860. https://doi.org/10.1145/2984511.2984565

[77] Eric Whitmire, Mohit Jain, Divye Jain, Greg Nelson, Ravi Karkar, Shwetak Patel, and Mayank Goel. 2017. DigiTouch: Reconfigurable Thumb-to-Finger Input and Text Entry on Head-mounted Displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–21. https://doi.org/10.1145/3130978

[78] Mathias Wilhelm, Daniel Krakowczyk, and Sahin Albayrak. 2020. Perisense: Ring-based multi-finger gesture interaction utilizing capacitive proximity sensing. *Sensors (Switzerland)* 20, 14 (2020), 1–23. https://doi.org/10.3390/s20143990

[79] Mathias Wilhelm, Daniel Krakowczyk, Frank Trollmann, and Sahin Albayrak. 2015. eRing: Multiple finger gesture recognition with one ring using an electric field. *ACM International Conference Proceeding Series* 25-26-June (2015), 1–6. https://doi.org/10.1145/2790044.2790047

[80] Erwin Wu, Ye Yuan, Hui Shyong Yeo, Aaron Quigley, Hideki Koike, and Kris M. Kitani. 2020. Back-hand-pose: 3D hand pose estimation for a wrist-worn camera via dorsum deformation network. *UIST 2020 - Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (2020), 1147–1160. https://doi.org/10.1145/3379337.3415897

[81] Xuhai Xu, Haitian Shi, Xin Yi, WenJia Liu, Yukang Yan, Yuanchun Shi, Alex Mariakakis, Jennifer Mankoff, and Anind K. Dey. 2020. *EarBuddy: Enabling On-Face Interaction via Wireless Earbuds*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376836

[82] Zheer Xu, Weihao Chen, Dongyang Zhao, Jiehui Luo, Te Yen Wu, Jun Gong, Sicheng Yin, Jialun Zhai, and Xing Dong Yang. 2020. BiTipText: Bimanual Eyes-Free Text Entry on a Fingertip Keyboard. *Conference on Human Factors in Computing Systems - Proceedings* (2020), 1–13. https://doi.org/10.1145/3313831.3376306

[83] Zheer Xu, Pui Chung Wong, Jun Gong, Te Yen Wu, Aditya Shekhar Nittala, Xiaojun Bi, Jürgen Steimle, Hongbo Fu, Kening Zhu, and Xing Dong Yang. 2019. TipText: Eyes-free text entry on a fingertip keyboard. *UIST 2019 - Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (2019), 883–889. https://doi.org/10.1145/3332165.3347865

[84] Yui Pan Yau, Lik Hang Lee, Zheng Li, Tristan Braud, Yi Hsuan Ho, and Pan Hui. 2020. How Subtle Can It Get? A Trimodal Study of Ring-sized Interfaces for One-Handed Drone Control. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020). https://doi.org/10.1145/3397319

[85] Sang Ho Yoon, Ke Huo, Vinh P. Nguyen, and Karthik Ramani. 2015. TIMMi: Finger-worn textile input device with multimodal sensing in mobile interaction. *TEI 2015 - Proceedings of the 9th International Conference on Tangible, Embedded, and Embodied Interaction* January (2015), 269–272. https://doi.org/10.1145/2677199.2680560

[86] Chaoyun Zhang, Paul Patras, and Hamed Haddadi. 2019. Deep learning in mobile and wireless networking: A survey. *IEEE Communications surveys & tutorials* 21, 3 (2019), 2224–2287.

[87] Cheng Zhang, Anandghan Waghmare, Pranav Kundra, Yiming Pu, Scott Gilliland, Thomas Ploetz, Thad E. Starner, Omer T. Inan, and Gregory D. Abowd. 2017. FingerSound: Recognizing unistroke thumb gestures using a ring. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–19. https://doi.org/10.1145/3130985

[88] Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Ruichen Meng, Sumeet Jain, Yizeng Han, Xinyu Li, Kenneth Cunefare, Thomas Ploetz, Thad Starner, Omer Inan, and Gregory D. Abowd. 2018. FingerPing: Recognizing fine-grained hand poses using active acoustic on-body sensing. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (2018), 1–10. https://doi.org/10.1145/3173574.3174011

[89] Xu Zhang, Xiang Chen, Wen-hui Wang, Ji-hai Yang, Lantz Vuokko, and Kong-qiao Wang. 2008. Hand Gesture Recognition and Virtual Game Control Based on 3D Accelerometer and EMG Sensors. *International Conference on Intelligent User Interfaces, Proceedings IUI* (2008), 1–5. papers3://publication/uuid/B648953B-F13C-4701-B0DF-93A63E4DEBB6

[90] Junhan Zhou, Yang Zhang, Gierad Laput, and Chris Harrison. 2016. AuraSense : Enabling Expressive Around - Smartwatch I nteractions with Electric Field Sensing. *UIST 2016 - Proceedings of the 29th Annual ACM Symposium on User Interface Software and Technology* Figure 1 (2016), 81–86.

[91] Kening Zhu, Simon Perrault, Taizhou Chen, Shaoyu Cai, and Roshan Lalintha Peiris. 2019. A sense of ice and fire: Exploring thermal feedback with multiple thermoelectric-cooling elements on a smart ring. *International Journal of Human Computer Studies* 130, February 2018 (2019), 234–247. https://doi.org/10.1016/j.ijhcs.2019.07.003

[92] Thomas G. Zimmerman, Joshua R. Smith, Joseph A. Paradiso, David Allport, and Neil Gershenfeld. 1995. Applying electric field sensing to human-computer interfaces. *Conference on Human Factors in Computing Systems - Proceedings* 1, May (1995), 280–287. https://doi.org/10.1145/223904.223940