EITPose: Wearable and Practical Electrical Impedance Tomography for Continuous Hand Pose Estimation

EITPose: Wearable and Practical Electrical Impedance Tomography for Continuous Hand Pose Estimation

Alexander Kyu*
akyu2@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Hongyu Mao*
hongyum@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Junyi Zhu
junyizhu@mit.edu
MIT CSAIL
Cambridge, Massachusetts, USA

Mayank Goel
mgoel1@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Karan Ahuja
kahuja@northwestern.edu
Northwestern University
Evanston, Illinois, USA

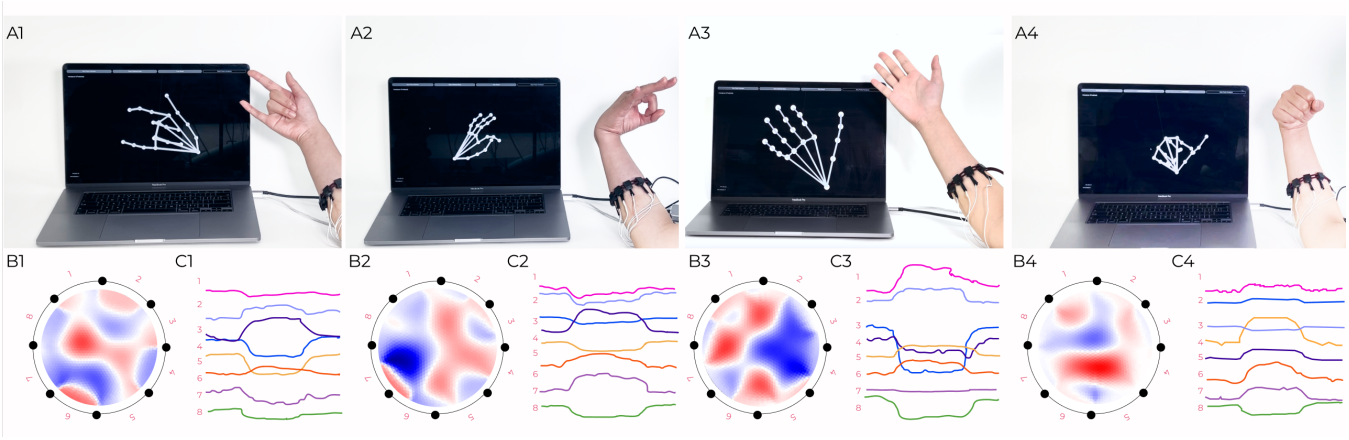

Figure 1: EITPose is a wrist-worn device for real-time hand pose estimation (A) that uses Electrical Impedance Tomography (EIT), to measure the impedance changes within the forearm. (B) shows a 2D reconstruction of this impedance map and (C) shows the average measured impedance signals per electrode emitter pair (1 through 8) while the user performs the hand pose.

ABSTRACT

Real-time hand pose estimation has a wide range of applications spanning gaming, robotics, and human-computer interaction. In this paper, we introduce EITPose, a wrist-worn, continuous 3D hand pose estimation approach that uses eight electrodes positioned around the forearm to model its interior impedance distribution during pose articulation. Unlike wrist-worn systems relying on cameras, EITPose has a slim profile (12 mm thick sensing strap) and is power-efficient (consuming only 0.3 W of power), making it an excellent candidate for integration into consumer electronic devices. In a user study involving 22 participants, EITPose achieves with a within-session mean per joint positional error of 11.06 mm. Its camera-free design prioritizes user privacy, yet it maintains cross-session and cross-user accuracy levels comparable to camera-based wrist-worn systems, thus making EITPose a promising technology for practical hand pose estimation.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; **Gestural input**; • **Computer systems organization** → *Sensors and actuators*; • **Hardware** → *Emerging interfaces.*

KEYWORDS

Electrical Impedance Tomography, Hand Pose, Hand Gesture, Input, Natural User Interfaces, Interaction Technique, Extended Reality

ACM Reference Format:
Alexander Kyu, Hongyu Mao, Junyi Zhu, Mayank Goel, and Karan Ahuja. 2024. EITPose: Wearable and Practical Electrical Impedance Tomography for Continuous Hand Pose Estimation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3613904.3642663

*Both authors contributed equally to this research.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642663

1 INTRODUCTION

Digitization of the hand has been a well-researched area for the last half a century and has applications in mixed reality [15], robotics [16], sports medicine [34], sign-language [19], spatial user interfaces [18], and more. Most successful and accurate implementations of hand-pose tracking involve direct sensing methods, mainly involving camera-based solutions. However, many of these direct sensing methods suffer from environmental noise and occlusion effects. Furthermore, camera-based methods raise privacy concerns, capturing unintended user and environmental data. Due to these issues, an accurate, privacy-sensitive, lightweight approach to digitization of the human hand remains unsolved.

To this end, we present EITPose, a new and practical wrist-worn device that uses Electrical Impedance Tomography (EIT) to continuously estimate hand pose in a mobile, self-contained, low-power and user-friendly form-factor (Figure 1). Tomography is a technique that produces detailed cross-sectional images of a medium's composition, using various forms of excitation such as electricity, sound waves, or radiation [5]. EIT uses pair-wise impedance measurements from surface electrodes surrounding the medium to recover its internal impedance distribution, making it particularly useful in medical contexts for imaging internal body functions in a non-invasive and affordable manner [9, 43].

Previous research has explored the use of wearable and mobile EIT form-factors for gesture classification [41, 42], activity detection [30], and rehabilitation [43]. However, utilizing a wrist-worn EIT band to accurately determine hand poses presents certain challenges. For instance, poor electrode connectivity can lead to added noise and impedance fluctuations due to factors such as the arm hair and air gaps. Furthermore, the human body presents a varied landscape for measurement, with individual differences in body composition and skin type further complicating the process. To ensure accurate hand pose estimation using EIT, it is crucial to establish a strategy that maintains stable electrode-to-skin connectivity and is robust to signal variations such as drift throughout an individual's session.

To this effect, EITPose extends upon EIT-Kit [44] (an open-source EIT toolkit) and showcases the feasibility of EIT for accurate and continuous hand pose estimation. It employs a device initialization step for optimal connectivity and signal-to-noise ratio, and proposes a novel waveform checker algorithm to ensure signal robustness throughout a user's session. Finally, EITPose implements additional user-friendly features, like enhancements on wireless capabilities and a band design that allows for easy adjustment across users and consistent electrode-to-skin contact.

To evaluate the efficacy of EITPose, we run user studies across a total of 22 participants to capture a diverse set of continuous hand poses across 12 different terminal poses and its corresponding EIT data. We conducted a series of evaluations across different scenarios. First, we evaluate the accuracy if the user had to calibrate the device for hand pose estimation every time they put on the device (within session) and assess performance consistency throughout that same session (even across long periods of use, spanning over an hour). Next, we evaluate hand pose estimation accuracy with a single calibration step (across sessions, with sessions collected a week apart to allow for more varied physiological changes). This is similar to how you might set up face ID on your iPhone when you first get your phone. We then evaluate EITPose's accuracy without individual per-user calibration (across users). This would be similar to an "out-of-the-box" experience where the device is already pre-trained and calibrated based on other users. Using an ensemble of trees machine learning model, EITPose had a mean per-joint hand pose estimation error of 11.06 mm, 17.81 mm, and 18.91 mm for within-session, cross-session and cross-user scenarios respectively. Lastly, we also benchmark EITPose's accuracy for hand gesture recognition (cross-user: 67.3%) to compare with prior works (Tomo [41], cross-user: 38.8%).

The contributions of this paper are multifold. Foremost, EITPose demonstrates the feasibility of continuous hand pose estimation using electrical impedance tomography. The improvements to sensor board, firmware, and band enables EITPose to achieve an accuracy comparative to camera-based methods while maintaining a low-power and privacy-preserving profile (Table 1). To enable future research to build on our approach and contribute to this domain, we will release our code, models, and the dataset upon publication.

2 RELATED WORK

There is a plethora of literature on wearable hand sensing. Prior works have primarily focused on hand gesture recognition, leveraging a diverse range of modalities such as infrared (IR) ranging [24], electromyography (EMG) [26], inertial measure unit (IMU) [37], acoustics [11, 12, 25], capacitive [2] and electrical impedance tomography [41, 42, 44] sensing. Refer to [6, 13, 23] for a full survey. In contrast, methods that capture the whole 3D hand pose offer a comprehensive and continuous sensing solution, which is more challenging and the focus of our work. Towards this end, researchers have made use of head-mounted cameras [1, 17, 27] or have instrumented the user's hand with gloves [8, 40] or MoCap markers[31, 35]. More relevant to this work, are approaches that instrument the wrist or forearm of the user, offering a practical, mobile and consumer friendly form-factor akin to the placement of smartwatch.

The predominant method for wrist-worn hand pose tracking is direct optical sensing, making use of a single camera [38] or an array[7, 10] of cameras. Commonly used optical sensors for hand pose estimation include RGB cameras [3, 36, 38], IR cameras [15, 29], and depth cameras [7, 32]. These camera-based systems usually make a trade-off between high field-of-view for sensing hands and practicality. Larger systems [3, 15] tend to be bulkier, but provide a clearer image of the hand, while slimmer more usable form-factors [36] suffer from occlusion artifacts. Moreover, the use of cameras encounters privacy concerns as these high-resolution sensors can inadvertently record unnecessary background details and sensitive information of both users and nearby individuals. To navigate these challenges, recent advances such as Discoband [7] and FingerTrak [10] employ an array of low-resolution depth sensors and thermal cameras respectively for hand pose tracking. These low-resolution cameras restrict data capture to essential hand shape details, substantially reducing privacy concerns. However, these optical systems consume a considerable amount of power. Furthermore, sensors must be positioned at a specific height above the arm to avoid occlusion caused by the wrist bending. Moreover,

Reference System	Sensing Modality	Privacy Preserving	Low Profile	Occlusion Invariant	Sampling Rate	Power Consumption	Within Session Error	Cross Session Error	Cross User Error
DiscoBand[7]	Depth Camera	✓	✓	✗	7.5Hz	3.6W	11.69mm	17.87mm	19.98mm
FingerTrak[10]	Thermal Camera	✓	✓	✗	16 Hz	High	12mm	~	87mm
Back-Hand-Pose[36]	RGB Camera	✗	✗	✗	~	High	8.81°*	9.77°*	9.72°*
Digits[15], Opisthenar[38]	IR Camera	✗	✗	✗	~	High	~	~	~
NeuroPose[21]	EMG	✓	✓	✓	40Hz	Low	6.24°*	~	~
EtherPose[14]	RF	✓	✗	✓	2.4Hz	4.5W	11.57mm	~	~
EITPose	EIT	✓	✓	✓	10Hz	0.3W	11.06 mm	17.81mm	18.91mm

Table 1: An overview of existing wrist-worn continuous hand pose estimation methods. All evaluations are mean per-joint positional error, apart from * which denotes mean per-joint angular error.

they are sensitive to environmental disruptions, including lighting and temperature interference, which can adversely affect tracking accuracy and reliability in outdoor settings.

In contrast, non-optical methods, although less extensively explored in hand pose tracking, emerge as a promising lightweight alternative to address privacy, invasiveness and occlusion concerns. These indirect sensing approaches measure features pertaining to pose without imaging the shape of the hand itself. For this, one of the most popular sensing modalities is Electromyography (EMG) which analyzes muscle activities for interpreting hand joints movements [20, 21]. Other approaches involve capacitive sensing [4] that reconstruct hand poses through capacitive images, or IMU's to predict hand orientation to extend the field-of-interaction for VR headsets [33]. However, these approaches faced limitations in terms of tracking accuracy and restricted tracking capabilities, thus hindering their practical implementations. In other recent works, Radio Frequency (RF) has been explored for hand digitization [14, 22]. Etherpose utilizes dual antennas to track continuous hand pose and nuanced micro-gestures by monitoring impedance changes correlating to alterations in the user's hand geometry. Despite a limited sampling rate and its large size, this method demonstrates advancements in non-optical hand pose tracking. While its accuracy is comparable to wrist-worn optical tracking methods when trained and tested within the same session, unlike them, it fails to generalize across sessions and different users.

Table 1 provides an overview of wrist-worn continuous hand pose estimation techniques. EITPose offers the robustness of optical tracking methods while maintaining the privacy and power profile of indirect sensing methods. It showcases strong within-session performance, generalizes across sessions (even after a week from the initial training), and achieves cross-user accuracies comparable to optical tracking approaches with a tenth of their power consumption.

3 IMPLEMENTATION

3.1 Sensing Principle

Electrical Impedance Tomography (EIT) is an imaging technique that uses surface electrodes to inject alternating currents (AC) and receive voltages to measure the impedance change of an object. The simplest configuration is a 2-pole setup [41] where a pair of electrodes injects an AC current at one electrode and reads the output signal at the other. In our 8 electrode setup, this would mean that we can measure the impedance across 28 different paths in the forearm, allowing us to determine the internal structure of the arm as well as the conductive properties of those internal structures. In a more complicated 4-pole setup [42], two adjacent electrodes inject the current source and the voltage is measured across two other adjacent electrodes. In our 8 electrode setup, this would allow us to measure the up to 40 different impedance paths through the forearm.

Our system includes an EIT sensing board with a 4-pole setup for measurements and an elastic wrist band made of eight stainless steel electrodes to ensure consistent contact across users. We build on EIT-kit's [44] open-source board design which incorporates a stable and robust current source to ensure the appropriate injection current and measurement gain to maximize the EIT signal while minimizing noise [44]. We extend this design for our application of hand pose estimation by adding extra filters in both hardware and software described below.

3.2 Sensing Board Design

Leveraging the foundational framework of the EIT-kit, the EITPose sensing board and firmware has undergone several refined modifications detailed below, enhancing its precision for hand pose estimation.

Device Initialization: Changes in underlying body factors, skin condition variations (sweat, dry skin, dead skin), and daily physiological fluctuations impair the utility of tomography techniques to generalize across sessions and users. To mitigate their effects, we perform an automated device initialization. Inspired by the auto-calibration feature of EITKit, we optimize the injection current and voltage measurement gain for each electrode pair to ensure connectivity and achieve an optimal signal-to-noise ratio every time the user wears the device.

Waveform Shape Checker: While a good device initialization ensures optimal connectivity at the start, as the session continues the signal can degrade and drift due to band movement and changing electrode-skin contact conditions during pose articulation. To counteract this we design a waveform shape checker algorithm wherein we calculate the mean absolute error between a measured sample of data points (for each measurement electrode pair) and an ideal sine wave that has been amplitude, phase and frequency normalized to the reference captured signal. If the mean absolute error is under our noise threshold, the signal passes the check and is used for pose inference. Else, we discard it and collect another sample of data points to compare against.

To ensure that the system doesn't get stuck trying to find a good sine wave in the scenario of bad contact, this waveform shape

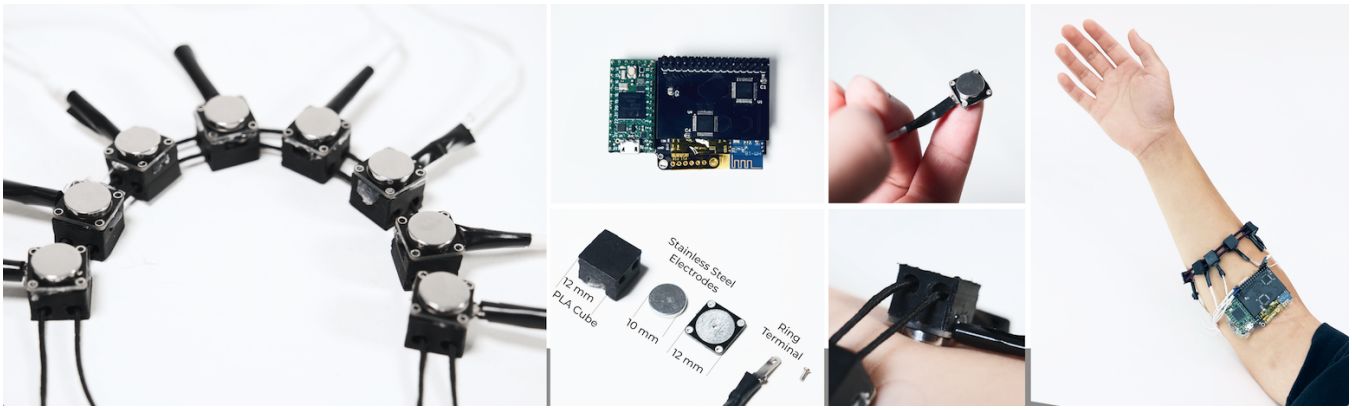

Figure 2: EITPose final wristband and sensing board. It includes a band with 8 stainless steel electrodes and the sensing PCB. Each stainless steel electrode is 10mm x 10mm x 2mm in size and are attached together by an elastic strap for easy adjustment. On the right is a picture of how the band is worn on a user.

checker only checks up to 10 times. If no signal passes the noise threshold till then, it returns a null signal (all 0's) signifying that the electrode pair is to be weighted the least during inference. Hence, it filters out noisy measurements during poor skin-to-electrode contact and allows for more consistent frame rates, while balancing optimal signal quality.

High-Pass Filter Implementation: EITPose augments the sensing board with an additional high-pass filtering circuit to mitigate the noise generated during MUX shifting, particularly noted during measurements conducted at frequencies below 3.68kHz. This addition is instrumental in delivering cleaner signals and more reliable data invariant to environmental noises.

Additional Features: We have upgraded the BLE (Bluetooth Low Energy) module on the EIT-kit to an HM-18 module, enhancing the board's wireless capabilities and data transfer capacity. Utilizing BLE 5.0 technology, it can now support a maximum frame rate of 28 FPS wirelessly in an 8-electrode setup. Additionally, we have integrated a BNO085 Inertial Measurement Unit (IMU) on board, which could be used for forearm orientation tracking in future work and help our system distinguish between wrist orientation and forearm orientation.

3.3 Band Design

Our wrist band was outfitted with an array of eight stainless electrodes to facilitate precise hand tracking (Figure 2). These cylindrical stainless-steel electrodes, featuring dimensions of 10mm x 10mm x 2mm and a slightly convex surface area of 78 mm^2 each, promise enhanced skin connectivity and longevity. To ensure precise placement and adaptability to varying forearm dimensions, the electrodes are individually mounted onto 3D-printed PLA cubes, with each cube measuring 12mm x 12mm x 8mm. These cubes are then connected to an elastic strap. This design ensures better electrode-skin contact and allows for flexible adjustment of electrode placements for varying forearm dimensions, allowing for even distribution around the forearm, facilitating accurate impedance measurements. In addition, these electrodes are interconnected through wires fortified by cold welding techniques, and are designed to interface with the pins on the MUX board of the customized EIT-kit sensing board.

In terms of performance, the device operates optimally within a sample rate range of 10 to 13 FPS for 8 electrodes, demonstrating stable signals at 10 FPS for the majority of participants. Furthermore, the device exhibits a marked reduction in power consumption, utilizing only 0.3W—tenfold lower than previous hand tracking methods like the DiscoBand and Etherpose — without compromising on the compactness of the band's design.

3.4 Data Processing and Machine Learning

Our input vector consists of a total of 48 signals. This includes 40 EIT signals (5 measurement pairs x 8 emitter electrode configurations), plus 8 additional signals generated from the averaging the RMS value for an electrode transmitter pair to each emitter pair. We demean each signal utilizing a rolling window of 40 seconds to eliminate signal shifting between different sessions and users, caused by both differing device initialization parameters and measurement changes due to differing impedance's from physiological differences between users. Note, that the window only looks at past frames and makes use of no future data - in scenarios when the data available is less than 40s, it makes use of the existing available data in the window.

This demeaned input vector of 48 is the input to our machine learning model, for which we employ SciPy's ExtraTreesRegressor (default parameters, 100 estimators). We chose this ensemble of trees machine learning method due to its robustness to missing data, invariance to scaling, and outliers. The output of our model is a vector of size 63, representing 21 3D hand pose key points. Similarly, for our gesture classifier we employ a SciPy's ExtraTreesClassifier (default parameters, 100 estimators) that takes the demeaned input vector and outputs a gesture class.

4 DATA COLLECTION PROCEDURE

To evaluate the effectiveness of the EITPose system, we recruited 22 participants (11 male, 11 female, mean age of 26 years) across two user study protocols: Within-session longitudinal study (3

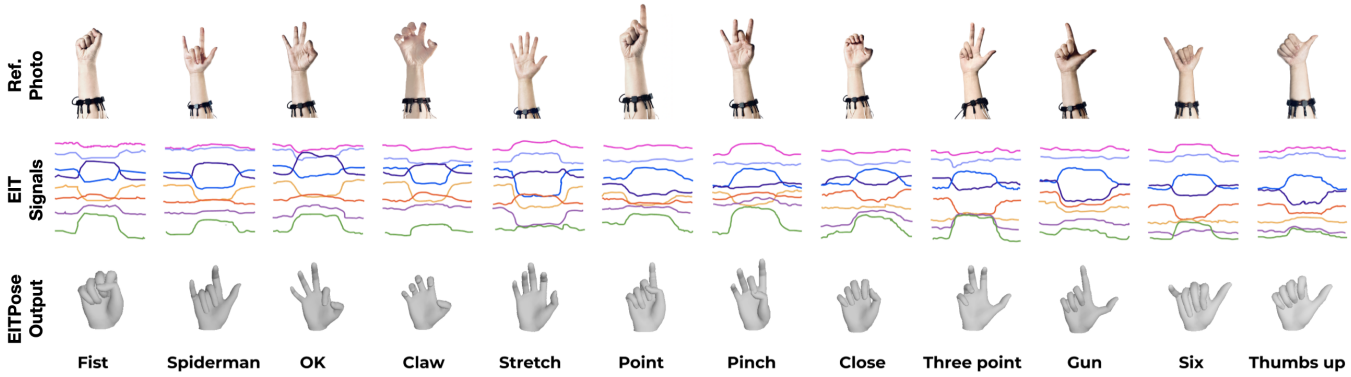

Figure 3: Top row: Reference images of the 12 terminal hand poses used in our study. Note, these were used as destination poses and all intermediate pose states were continuously captured providing a diverse set. Middle row: visualizer depicting the 8 average measured impedance signals for each electrode emitter pair while the user performs the hand pose. Bottom row: Predicted hand poses, visualized using a Mano hand mesh model.

participants) and Cross-session & Cross-user study (19 participants). Each protocol differs in the number and spacing of data collection sessions. The heights of the participants ranged from 1.52 to 1.86 m (mean = 1.70 m), and forearm circumference varied from 20 to 31 cm (mean = 24.96 cm). The participants were compensated with 15 USD for their participation. The data collection apparatus consisted of a GUSGU G910 camera positioned at a 45-degree at one end to capture ground truth hand pose against a white background. Akin to prior works like Etherpose and Discoband, we make use of Google's MediaPipe [39] sampling at 20 Hz (2560 by 1440 image resolution) to estimate ground truth hand pose (21 3D hand key points with the palm as the origin). We further scale the bone lengths according to the MANO hand model [28] to estimate mesh parameters and produce anatomically coherent hands (Figure 3).

For each session, participants were seated at a table and were fitted with the band, arranging the first electrode near the middle of their inner forearm and placing the remaining electrodes in a clockwise pattern around it. We did not mark fixed positions for the electrodes to introduce more variability into the data. We then performed device initialization (Section 3.2) and adjusted the elastic strap and aligned the position of the electrodes to ensure optimal connectivity and comfort. Upon briefing the participants about the study they were instructed to extend their arms and hands in front of the camera alongside a computer monitor, which provided visual instructions.

For data collection, we designed twelve common hand poses drawn from prior works (Figure 3). Participants were prompted to perform the poses as shown on the computer display. These acted as terminal poses and participants were asked to naturally shift between the different poses and even change wrist orientations between them, allowing us to capture a diverse set of intermediate pose states. When a new frame of data arrives from our prototype (at 10 Hz), the most recent MediaPipe hand keypoints are recorded along side the EIT data. A single round of data collection consisted of a random ordering of the twelve hand poses. We collected eight rounds of hand pose data in this fashion, which formed one trial of data collection (lasting for roughly 10 mins for each participant).

4.1 Within-session Longitudinal Data Collection

Most prior works [7, 14, 41] make use of data collected over back to back sessions, with each session only lasting less than 10 mins. This fails to account for the within-session variability typically depicted in the real-world that consumer devices have to contend with (e.g. VR gaming session lasting an hour). To this effect, we recruited 3 participants to evaluate the within-session accuracy of EITPose over continued periods of use (spanning about 1.75 hours for each participant). This data collection consisted of a single session comprising of 4 trials. The session started with a data collection trial, followed by a 20-minute gap. This process continued with 3 more data collection trials, each separated by 20 minute gaps, resulting altogether in 4 trials and 3 20-minute gaps between them. Note, there was a single device initialization at the beginning of the session (before the first trial) and the device was not taken off or re-initialized for the subsequent trials.

In the 20 mins gap between sessions, participants engaged in activities of their own choosing with the band on, ensuring that the data was streaming continuously during the whole time. No ground truth data was collected during this time as the participants were free to move around. The activities performed by the participants ranged from doing homework on a laptop (including typing, using a mouse), eating food, drinking water, talking to people in their vicinity and using their phones. At the end of each gap, participants returned to the apparatus to start the next data collection trial.

4.2 Cross-session & Cross-user Data Collection

Here we recruited 19 participants to evaluate the cross-user and cross-session efficacy of EITPose. All participants completed a first data collection session consisting of two back to back trials. The participants were asked to return a week later for a second data collection session. This week-long time window allowed for more varied physiological changes (skin conditions such as dryness or sweat, weight fluctuations, alterations in muscle and tissue, etc.) to manifest in comparison to prior works [14, 41] which collect cross-session data within the same hour. This helps us evaluate

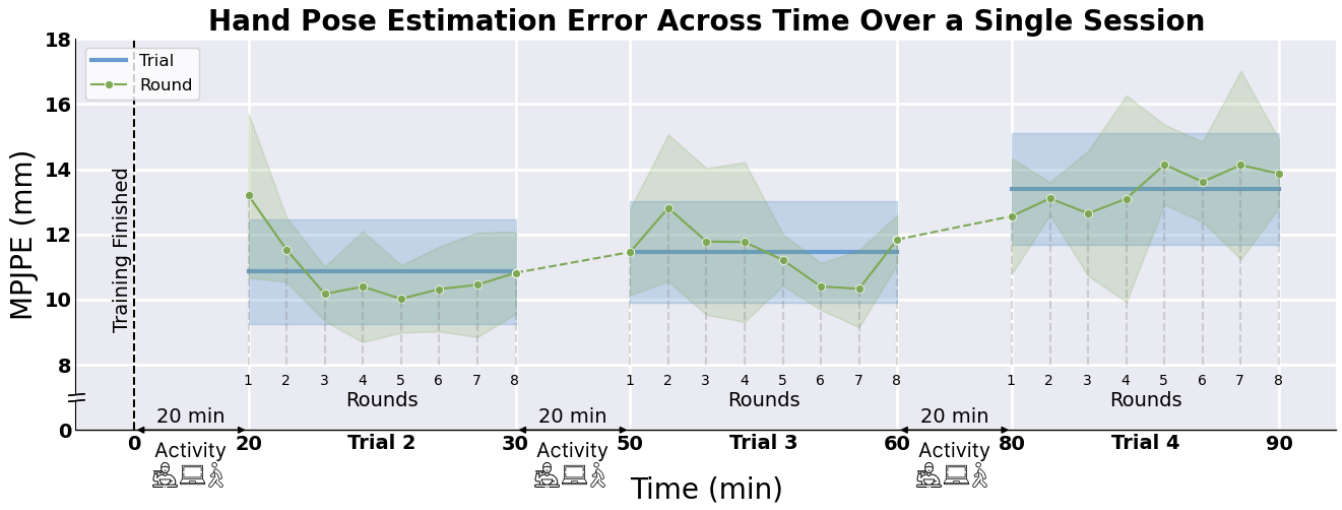

Figure 4: EITPose mean per-joint positional error (MPJPE) averaged over participants across time for a longitudinal session. Filled Areas/Error bars indicate standard deviation.

the robustness of EITPose on a more varied and potentially noisier cross-session data.

10 of the 19 participants returned for the second data collection session. Session I consisted of two trials and Session II (collected one week later) consisted of a single trial, resulting in a total of three trials across the two sessions. Note, each session had a single device initialization at the beginning, that is the device was not taken off and re-initialized between the first two trials of Session I. Across all 19 participants, we collected a total of 380 mins of data (corresponding to approximately 210,000 data instances of hand poses with corresponding EIT data).

To further investigate hand gesture recognition accuracy, an expert annotator labelled the gestures across all trials corresponding to certain poses and hand orientations. In total 17 hand gestures were labeled corresponding to the 12 terminal poses (Figure 3): Fist, Spiderman, Ok, Claw, Stretch, Point, Pinch, Close, Three point, Gun, Six, Thumb up; and 5 additional poses: Relax, Up, Down, Left, Right to aid in comparison with prior works [41]. These additional poses were annotated from the start of the trial (Relax), and the different wrist pitch (Up & Down) and yaw (Left & Right) orientations of the Stretch hand pose. Each gesture corresponds to 10 EIT data points per round, resulting in 1360 data points per trial (1 trial = 17 gestures × 10 data points × 8 rounds).

5 RESULTS

We designed our evaluation procedure in order to analyze and isolate different factors that affect performance. First, we want to investigate EITPose's feasibility to detect continuous hand poses throughout a continued longitudinal session across different hand configurations and motions (Section 5.1). Further, we also want to quantify its robustness for a single user across multiple worn sessions (Section 5.2), and across multiple users (Section 5.3). This helps us analyze the efficacy of our system to varied real-life conditions, including changes in skin conditions, muscle tone, and electrode placements and Body Mass Index (BMI). To evaluate pose accuracy we compute metrics such as mean per-joint positional error (MPJPE) between the predicted and ground truth 3D hand pose. Table 1 provides an overview of our accuracy in comparison to prior direct (optical) sensing and indirect (non-optical) sensing methods. Lastly, we evaluate hand gesture recognition accuracy (Section 5.4) to benchmark EITPose with prior works.

5.1 Within Session Performance

We conducted an investigation of within session evaluation to assess the performance of EITPose in scenarios where the user might train or calibrate the sensor each time its worn. First, we evaluate its performance over periods of continued use (spanning over an hour and a half) making use of our Longitudinal data collected on 3 participants (Section 4.1) across 4 trials of a single session. For this, we trained our hand pose regression model on Trial 1 (8 rounds, 12 terminal hand poses) and tested it on Trial 2, 3 and 4. Figure 4 shows the results averaged across all participants for each round and trial. Over time, we do notice an increase in MPJPE, increasing from 10.87 mm in Trial 2, to 11.45 mm for Trial 3 and subsequently to 13.40 mm for Trial 4.

In addition to a longitudinal single session evaluation, we also want to test the generalizability of EITPose's within-session performance over a wider participant pool and demographic. For this we make use of our data collection over 19 participants (Section 4.2). Here, we trained our pose regression model on one trial (8 rounds, 12 terminal hand poses) of session I and tested it on the other holdout trial. This resulted in two train/test combinations from the two trials within Session I. The averaged results across both combinations across all 19 participants gives a MPJPE of 10.93 mm (SD=4.69 mm). These results are broken down by hand joints in Figure 5 (with the palm as the origin). We note that the error accumulates along the hand pose kinematic chain - the average error of the end-effectors (fingertips and thumb) is 16.82 mm. This is

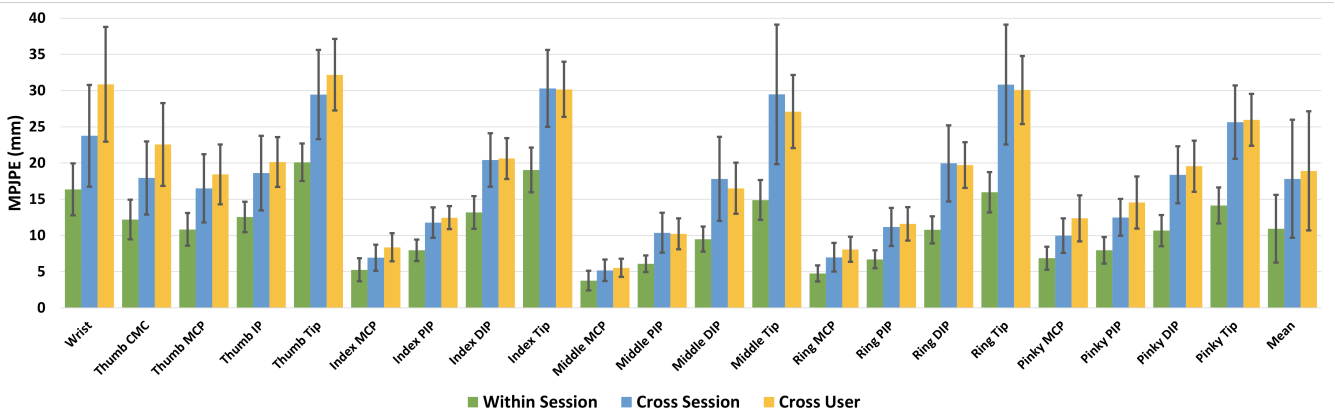

Figure 5: EITPose mean per-joint positional error (MPJPE) for hand pose estimation averaged across all participants. Error bars indicate standard deviation.

expected as these joints exhibit higher ranges of motion and it's harder to indirectly sense the joints farther away from the point of instrumentation. Further, the wrist exhibits a larger than mean error of 16.36 mm which can be attributed to the diverse range of pose orientations it has to contend with.

Across all 22 participants EITPose has a within-session MPJPE of 11.06 mm. This is comparable to the performance of prior works such as DiscoBand (depth camera based), FingerTrak (thermal camera based) and EtherPose (RF based) which have a similar error profiles of 11.69, 12.00 and 11.57 mm respectively.

5.2 Cross Session Performance

In addition to evaluating performance within a session, it is important for a sensing system to maintain performance each time a user wears the device, without requiring any retraining. Notably, it is key to evaluate this not just after a training session, but some time in the future to assess changes in placement, skin conditions and underlying muscle tone among others. To assess this, we performed a leave-one-session-out cross-validation evaluation, between session I and session II data (captured a week apart) across the 10 participants from our Cross-session data collection (Section 4.2). Here, we initially trained the model on data from session I and tested it on session II, and vice versa. The results from these evaluations were averaged to derive the performance metrics for a single user. This procedure was repeated for all users and results aggregated, as illustrated in Figure 5. The average MPJPE of EITPose across different sessions is 17.81 mm (SD=8.15). This is in line with prior optical methods such as DiscoBand which have a cross-session error of 17.87 mm.

5.3 Cross User Performance

In order to evaluate the "out-of-the-box" experience of EITPose, we perform a cross-user performance evaluation across our 19 participants from our Cross-user data collection (Section 4.2). This is akin to a pre-configured model that is trained once with a corpus of data but works as is, without needing any data from the target user. To explore this aspect, we conducted a leave-one-user-out cross-validation, wherein data from 18 participants were utilized for training, and the 19th participant's data was reserved for testing. We then average the results across all users. For cross-user continuous hand pose tracking, we found a MPJPE of 18.91 mm (SD=8.22), broken down per joint in Figure 5. Akin to within-session error profiles, the error increases as we move away from the palm, being higher for the fingertips. The accuracy of EITPose is in line with DiscoBand which has an error of 19.98 mm but a power consumption of 3.6 W while that of our system is 0.3 W.

5.4 Hand Gesture Recognition

In order to benchmark against prior hand gesture recognition works, we also evaluate our systems accuracy along such lines. We make use of our gesture annotated 19 participant data described in Section 4.2 for evaluation. Across all 17 hand gestures, following a leave-one-trial out cross-validation protocol, EITPose has a within-session gesture recognition accuracy of 81.69% (SD = 8.18%). Employing a leave-one-user out cross-validation protocol, it has a cross-user accuracy of 40.52% (SD = 11.39%) across all 17 gestures.

We also take a subset of our 17 gestures to match the 8-class gesture set of Tomo [41], which employs EIT for hand gesture classification. Figure 6 provides the confusion matrices of EITPose for both within-session and cross-user evaluation protocols on the 8-class Tomo hand gesture set. Across all 8 gestures, EITPose has an accuracy of 93.8% and 67.3% for within-session and cross-user evaluations, respectively. In comparison to wrist-based Tomo's 96.6% (within-session) and 38.8% (cross-user), EITPose has a similar within-session performance, but significantly outperforms Tomo across users. As another comparison point, BeamBand [11] (a wrist-worn band utilizing acoustic sensing), has a within-session accuracy of 92.5% (vs 93.8% of EITPose) cross-user gesture accuracy of 51.7% (vs 67.3% of EITPose) on the Tomo set.

6 OPEN SOURCE

To enable other researchers and practitioners to build upon our work, we will make our processing pipeline, trained models, dataset and annotations available at https://github.com/SPICExLAB/EITPose

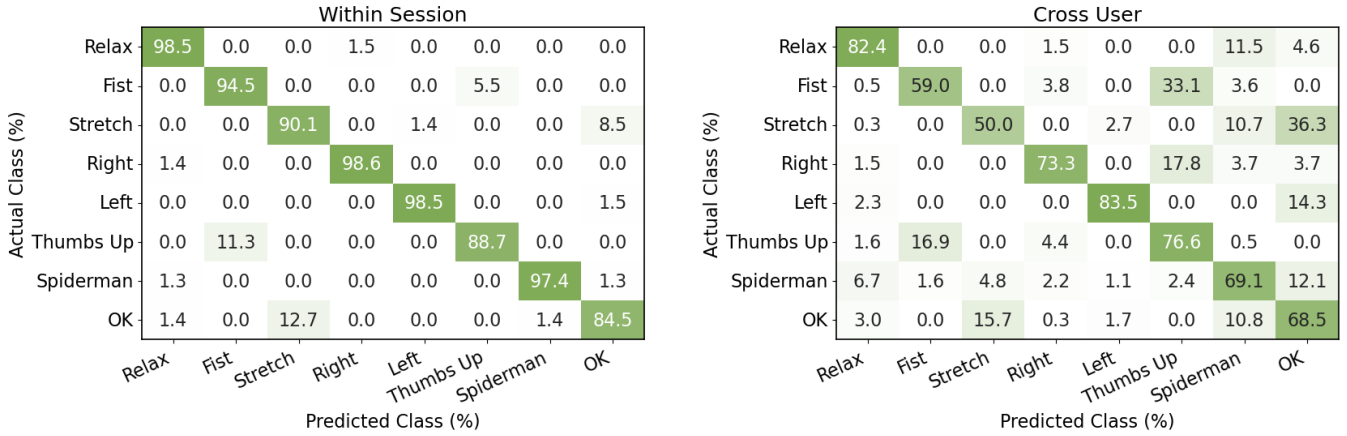

Figure 6: EITPose confusion matrices on Tomo[41] hand gesture set for Within-Session (left) and Cross-User (right) evaluation.

with the gracious permission of our participants. This will help facilitate replication and further exploration in the field.

7 LIMITATIONS & FUTURE WORK

While EITPose exhibits great potential as a generalizable hand pose estimation technique, it is essential to underscore several significant drawbacks, which, in turn, provide valuable insights into potential areas for future research. First, there is still a significant drop in accuracy going from within-session (11.06 mm) performance to cross-session (17.81 mm) and cross-user (18.91 mm) performance. This is common in optical wrist based sensing techniques as well. We were, however, surprised to see that the error for our cross-user evaluation only increased a little (by 6%) from our cross-session evaluation. Thus in the future, going for a big data approach of training a data-intensive global machine learning model and tuning it per user might be a viable exploration to further improve accuracy.

We also note that our current sensing armband has limited spatial resolution. EITPose makes use of 8 electrodes with a total of 40 sensor-pair values. EIT-kit [44] is capable of utilizing up to 64 electrodes which would substantially increase the number of sensor-pairs to 3904 measurements and thus the spatial impedance resolution. Works by Zhang et al. have shown a hand gesture recognition accuracy improvement of 5.3% when going from a 8 to 32 electrode setup. However, this comes at a cost of sampling framerate, decreasing it to 3Hz for a 32 electrode setup. In the future, this trade-off should be evaluated, especially in this context of real-time hand pose tracking.

Lastly, while our band has the advantage of being relatively unaffected by environmental noises like light, heat, or occlusion effects due to the nature of the sensing modality, the effects of other factors like skin contact conditions and sweat may affect the EIT signal. Furthermore, different electrode placements might result in a completely different signal profile than one used during training. In such scenarios, in the future, we would like to integrate the orientation from our on-board IMU to better model how the EIT signals shift with changes in band rotation and auto-calibrate based on it.

8 CONCLUSION

We have presented EITPose, a mobile and self-contained form-factor arm band that enables continuous 3D hand pose estimation using Electrical Impedance Tomography (EIT) to capture the interior impedance geometry of a user's arm. Our approach allows for a minimally invasive (12 mm) and power-efficient (0.3 W) form factor, showcasing the potential for integration into consumer electronic form factors. In a user study across 22 participants, EITPose has a within-session mean per joint error of 11.06 mm. Further, its cameraless approach enables a privacy-sensitive sensing system but with the cross-session and cross-user accuracy profiles of camera-based wrist worn systems, thus providing the best of both worlds.

REFERENCES

[1] Karan Ahuja, Chris Harrison, Mayank Goel, and Robert Xiao. 2019. Mecap: Whole-body digitization for low-cost vr/ar headsets. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 453–462.

[2] Karan Ahuja, Paul Streli, and Christian Holz. 2021. TouchPose: hand pose prediction, depth estimation, and touch classification from capacitive images. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 997–1009.

[3] Riku Arakawa, Azumi Maekawa, Zendai Kashino, and Masahiko Inami. 2020. Hand with Sensing Sphere: Body-Centered Spatial Interactions with a Hand-Worn Spherical Camera. In *Proceedings of the 2020 ACM Symposium on Spatial User Interaction* (Virtual Event, Canada) *(SUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 1, 10 pages. https://doi.org/10.1145/3385959.3418450

[4] Kazuyuki Arimatsu and Hideki Mori. 2020. Evaluation of Machine Learning Techniques for Hand Pose Estimation on Handheld Device with Proximity Sensor. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376712

[5] MS Beck, T Dyakowski, and RA Williams. 1998. Process tomography-the state of the art. *Transactions of the Institute of Measurement and Control* 20, 4 (1998), 163–177.

[6] Weiya Chen, Chenchen Yu, Chenyu Tu, Zehua Lyu, Jing Tang, Shiqi Ou, Yan Fu, and Zhidong Xue. 2020. A Survey on Hand Pose Estimation with Wearable Sensors and Computer-Vision-Based Methods. *Sensors* 20, 4 (2020). https://doi.org/10.3390/s20041074

[7] Nathan Devrio and Chris Harrison. 2022. DiscoBand: Multiview Depth-Sensing Smartwatch Strap for Hand, Body and Environment Tracking. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 56, 13 pages. https://doi.org/10.1145/3526113.3545634

[8] Oliver Glauser, Shihao Wu, Daniele Panozzo, Otmar Hilliges, and Olga Sorkine-Hornung. 2019. Interactive hand pose estimation using a stretch-sensing soft glove. *ACM Transactions on Graphics (ToG)* 38, 4 (2019), 1–15.

[9] David S Holder. 2004. *Electrical impedance tomography: methods, history and applications.* CRC Press.

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

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

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

[13] Shuo Jiang, Peiqi Kang, Xinyu Song, Benny P.L. Lo, and Peter B. Shull. 2022. Emerging Wearable Interfaces and Algorithms for Hand Gesture Recognition: A Survey. *IEEE Reviews in Biomedical Engineering* 15 (2022), 85–102. https://doi.org/10.1109/RBME.2021.3078190

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

[15] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: Freehand 3D Interactions Anywhere Using a Wrist-Worn Gloveless Sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (Cambridge, Massachusetts, USA) *(UIST '12)*. Association for Computing Machinery, New York, NY, USA, 167–176. https://doi.org/10.1145/2380116.2380139

[16] Tomáš Kot and Petr Novák. 2018. Application of virtual reality in teleoperation of the military mobile robotic system TAROS. *International Journal of Advanced Robotic Systems* 15, 1 (2018), 1729881417751545. https://doi.org/10.1177/1729881417751545 arXiv:https://doi.org/10.1177/1729881417751545

[17] UltraLeap 2023. *UltraLeap LeapMotion.* UltraLeap. https://docs.ultraleap.com/

[18] Jinha Lee, Alex Olwal, Hiroshi Ishii, and Cati Boulanger. 2013. SpaceTop: Integrating 2D and Spatial 3D Interactions in a See-through Desktop Environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 189–192. https://doi.org/10.1145/2470654.2470680

[19] Yun Li, Xiang Chen, Jianxun Tian, Xu Zhang, Kongqiao Wang, and Jihai Yang. 2010. Automatic Recognition of Sign Language Subwords Based on Portable Accelerometer and EMG Sensors. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction* (Beijing, China) *(ICMI-MLMI '10)*. Association for Computing Machinery, New York, NY, USA, Article 17, 7 pages. https://doi.org/10.1145/1891903.1891926

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

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

[22] Yilin Liu, Shijia Zhang, Mahanth Gowda, and Srihari Nelakuditi. 2022. Leveraging the Properties of MmWave Signals for 3D Finger Motion Tracking for Interactive IoT Applications. *Proc. ACM Meas. Anal. Comput. Syst.* 6, 3, Article 52 (dec 2022), 28 pages. https://doi.org/10.1145/3570613

[23] Jess McIntosh. 2019. *Exploring the practicality of wearable gesture recognition.* PhD thesis. The University of Bristol, Bristol, England. Available at https://research-information.bris.ac.uk/files/187498382/Final_Copy_2019_01_23_McIntosh_J_PhD_Redacted.pdf.

[24] Jess McIntosh, Asier Marzo, and Mike Fraser. 2017. Sensir: Detecting hand gestures with a wearable bracelet using infrared transmission and reflection. In *Proceedings of the 30th annual ACM symposium on user interface software and technology*. 593–597.

[25] Jess McIntosh, Asier Marzo, Mike Fraser, and Carol Phillips. 2017. EchoFlex: Hand Gesture Recognition Using Ultrasound Imaging. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1923–1934. https://doi.org/10.1145/3025453.3025807

[26] Thalmic Labs 2022. *Myo EMG armband.* Thalmic Labs. https://developerblog.myo.com/

[27] Meta 2023. *Oculus Handtracking.* Meta. https://developer.oculus.com/documentation/unity/unity-handtracking/

[28] Javier Romero, Dimitrios Tzionas, and Michael J. Black. 2017. Embodied Hands: Modeling and Capturing Hands and Bodies Together. *ACM Trans. Graph.* 36, 6, Article 245 (nov 2017), 17 pages. https://doi.org/10.1145/3130800.3130883

[29] Farshid Salemi Parizi, Wolf Kienzle, Eric Whitmire, Aakar Gupta, and Hrvoje Benko. 2021. RotoWrist: Continuous Infrared Wrist Angle Tracking Using a Wristband. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (Osaka, Japan) *(VRST '21)*. Association for Computing Machinery, New York, NY, USA, Article 26, 11 pages. https://doi.org/10.1145/3489849.3489886

[30] Munehiko Sato, Rohan S. Puri, Alex Olwal, Yosuke Ushigome, Lukas Franciszkiewicz, Deepak Chandra, Ivan Poupyrev, and Ramesh Raskar. 2017. *Zensei: Embedded, Multi-Electrode Bioimpedance Sensing for Implicit, Ubiquitous User Recognition.* Association for Computing Machinery, New York, NY, USA, 3972–3985. https://doi.org/10.1145/3025453.3025536

[31] Matthias Schröder, Jonathan Maycock, and Mario Botsch. 2015. Reduced Marker Layouts for Optical Motion Capture of Hands. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games* (Paris, France) *(MIG '15)*. Association for Computing Machinery, New York, NY, USA, 7–16. https://doi.org/10.1145/2822013.2822026

[32] Srinath Sridhar, Anders Markussen, Antti Oulasvirta, Christian Theobalt, and Sebastian Boring. 2017. WatchSense: On- and Above-Skin Input Sensing through a Wearable Depth Sensor. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3891–3902. https://doi.org/10.1145/3025453.3026005

[33] Paul Streli, Rayan Armani, Yi Fei Cheng, and Christian Holz. 2023. HOOV: Hand Out-Of-View Tracking for Proprioceptive Interaction using Inertial Sensing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 310, 16 pages. https://doi.org/10.1145/3544548.3581468

[34] Evan A. Susanto, Raymond KY Tong, Corinna Ockenfeld, and Newmen SK Ho. 2015. Efficacy of robot-assisted fingers training in chronic stroke survivors: a pilot randomized-controlled trial. *Journal of NeuroEngineering and Rehabilitation* 12, 42 (2015). https://doi.org/10.1186/s12984-015-0033-5 arXiv:https://doi.org/10.1186/s12984-015-0033-5

[35] vicon 2022. *vicon.* vicon. https://www.vicon.com/

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

[37] Xuhai Xu, Jun Gong, Carolina Brum, Lilian Liang, Bongsoo Suh, Shivam Kumar Gupta, Yash Agarwal, Laurence Lindsey, Runchang Kang, Behrooz Shahsavari, et al. 2022. Enabling hand gesture customization on wrist-worn devices. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.

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

[39] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. MediaPipe Hands: On-device Real-time Hand Tracking. *CoRR* abs/2006.10214 (2020). arXiv:2006.10214 https://arxiv.org/abs/2006.10214

[40] Xiao Zhang, Griffin Klevering, Juexing Wang, Li Xiao, and Tianxing Li. 2023. RoFin: 3D Hand Pose Reconstructing via 2D Rolling Fingertips. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services* (Helsinki, Finland) *(MobiSys '23)*. Association for Computing Machinery, New York, NY, USA, 330–342. https://doi.org/10.1145/3581791.3596838

[41] Yang Zhang and Chris Harrison. 2015. Tomo: Wearable, Low-Cost Electrical Impedance Tomography for Hand Gesture Recognition. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 167–173. https://doi.org/10.1145/2807442.2807480

[42] Yang Zhang, Robert Xiao, and Chris Harrison. 2016. Advancing Hand Gesture Recognition with High Resolution Electrical Impedance Tomography. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 843–850. https://doi.org/10.1145/2984511.2984574

[43] Junyi Zhu, Yuxuan Lei, Aashini Shah, Gila Schein, Hamid Ghaednia, Joseph Schwab, Casper Harteveld, and Stefanie Mueller. 2022. MuscleRehab: Improving Unsupervised Physical Rehabilitation by Monitoring and Visualizing Muscle Engagement. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 33, 14 pages. https://doi.org/10.1145/3526113.3545705

[44] Junyi Zhu, Jackson C Snowden, Joshua Verdejo, Emily Chen, Paul Zhang, Hamid Ghaednia, Joseph H Schwab, and Stefanie Mueller. 2021. EIT-Kit: An Electrical Impedance Tomography Toolkit for Health and Motion Sensing. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 400–413. https://doi.org/10.1145/3472749.3474758