Figure 1: We use a capacitive sensing technique to measure changes in wrist profiles (A), which allows inferring the force (B) and grasp (C) used to interact with external objects. This enables digital input on unaugmented physical objects mediated by a wristband (D).

ABSTRACT

We demonstrate rich inferences about unaugmented everyday objects and hand object interactions by measuring minute skin surface deformations at the wrist using a sensing technique based on capacitance. The wristband prototype infers muscle and tendon tension, pose, and motion, which we then map to force (9 users, 13.66 +/- 9.84 N regression error on classes 0–49.1 N), grasp (9 users, 81 +/- 7 % classification accuracy on 6 grasps), and continuous interaction (10 users, 99 +/- 1 % discrimination accuracy between 6 interactions, 89–97 % accuracy on 3 states within each interaction) using basic machine learning models.

We wrapped these sensing capabilities into a proof-of-concept end-to-end system, Ubiquitous Controls, that enables virtual range inputs by sensing continuous interactions with unaugmented objects. Eight users leveraged our system to control UI widgets (like sliders and dials) with object interactions (like “cutting with scissors” and “squeezing a ball”). Finally, we discuss the implications and opportunities of using hands as a ubiquitous sensor of our surroundings.

CCS CONCEPTS

• Human-centered computing → Interaction techniques; Ubiquitous and mobile computing systems and tools; Mixed / augmented reality; Interaction devices.

KEYWORDS

wrist topography, capacitive sensing, wristband, everyday objects, affordances

ACM Reference Format:

1 INTRODUCTION

As a wearable device on the wrist, smartwatches offer a useful vantage point to sense the prehensile movements of the hand. Recent research has demonstrated how sensing the fine-grained hand activity provides context about user activity and intent, and even object recognition. Prior work has demonstrated classifying repeated, fine hand motions via acceleration signals [40], as well as identifying objects held in the hand via EMG signals [17]. In this work, we examine how we can further increase contextual awareness about user activities and their environment: we infer properties of unaugmented everyday objects and the user’s interaction with those objects by measuring continuous, instantaneous changes of wrist topography with capacitive signals.

Our wristband prototype infuses an electrical signal into the body and houses a capacitive sensor matrix to measure topographical changes at the wrist (see Figure 1). As the distance between a localized patch of skin and a sensor’s receiver antenna varies, so does the capacitive charge. By measuring these minute skin surface deformations at multiple points near the wrist, it is possible to infer isometric muscle tension, tendon movement, and hand pose and motion.

Using machine learning, we map anatomical features like tendon tension, pose, and movement beneath the skin’s surface to interaction properties like force, grasp, and object shape. Additionally, our approach enables discrimination of hand gestures and extends previous gesture recognition research by also precisely regressing the state within interaction ranges of unaugmented objects and the user’s interaction with those objects by measuring continuous, instantaneous changes of wrist topography with capacitive signals.

Our wristband prototype infuses an electrical signal into the body and houses a capacitive sensor matrix to measure topographical changes at the wrist (see Figure 1). As the distance between a localized patch of skin and a sensor’s receiver antenna varies, so does the capacitive charge. By measuring these minute skin surface deformations at multiple points near the wrist, it is possible to infer isometric muscle tension, tendon movement, and hand pose and motion.

Using machine learning, we map anatomical features like tendon tension, pose, and movement beneath the skin’s surface to interaction properties like force, grasp, and object shape. Additionally, our approach enables discrimination of hand gestures and extends previous gesture recognition research by also precisely regressing the state within interaction ranges of unaugmented objects and the user’s interaction with those objects by measuring continuous, instantaneous changes of wrist topography with capacitive signals.

Our wristband prototype infuses an electrical signal into the body and houses a capacitive sensor matrix to measure topographical changes at the wrist (see Figure 1). As the distance between a localized patch of skin and a sensor’s receiver antenna varies, so does the capacitive charge. By measuring these minute skin surface deformations at multiple points near the wrist, it is possible to infer isometric muscle tension, tendon movement, and hand pose and motion.

Using machine learning, we map anatomical features like tendon tension, pose, and movement beneath the skin’s surface to interaction properties like force, grasp, and object shape. Additionally, our approach enables discrimination of hand gestures and extends previous gesture recognition research by also precisely regressing the state within interaction ranges of unaugmented objects and the user’s interaction with those objects by measuring continuous, instantaneous changes of wrist topography with capacitive signals.

Our wristband prototype infuses an electrical signal into the body and houses a capacitive sensor matrix to measure topographical changes at the wrist (see Figure 1). As the distance between a localized patch of skin and a sensor’s receiver antenna varies, so does the capacitive charge. By measuring these minute skin surface deformations at multiple points near the wrist, it is possible to infer isometric muscle tension, tendon movement, and hand pose and motion.

Using machine learning, we map anatomical features like tendon tension, pose, and movement beneath the skin’s surface to interaction properties like force, grasp, and object shape. Additionally, our approach enables discrimination of hand gestures and extends previous gesture recognition research by also precisely regressing the state within interaction ranges of unaugmented objects and the user’s interaction with those objects by measuring continuous, instantaneous changes of wrist topography with capacitive signals.

In the case of user wearables, we can acquire information about an individual’s activity and recognize objects in the user’s environment, e.g., with electromagnetic signals [42], EMG [17], bio-acoustic signals [41], or audio plus inertial data [40]. This monitoring allows a portable system to display contextual information, but offers limited dynamic control of interactive systems. Others explore different sensing techniques to detect freestanding gestures: Digits [37] leverages an infrared camera, Wristflex [15] uses an array of force sensitive resistors, Tomo [83] relies on electrical impedance tomography, Serendipity [76] and TapID [54] employ IMUs, SensIR [52] uses near-infrared sensing, and Capband [70] explores flexible capacitive sensors. Others use EMG [55], but given muscle locations this is more suited to the forearm [35]. Interferi [33] measures ultrasonic signals with piezoelectric transducers, and beyond recognizing gestures investigates continuous tracking tasks like regressing weights and smile intensity. Our work goes in a similar direction by focusing on properties of the interaction between hand and object, and additionally uses more-granular activity monitoring to precisely map pose and interaction to virtual controls. Another wearable system focused on properties of hand-object interaction is Fsense [6], which uses photoplethysmography to discriminate two levels of force during common gestures.

While we also target a wristband form factor, we leverage wrist topography (broadly related to mechanomyography or MMG, first described by Grimaldi in 1665 [24]) to measure motions and hand activity. With MMG, an accelerometer, microphone, or high-accuracy laser near a muscle can “hear” its motion, and even infer fatigue [69]. MMG has been deeply explored in medical literature (for a review see [38] or [34]), but has seen little adoption in HCl in spite of the fact that it has been shown to have a higher signal-to-noise ratio than electromyography (EMG) [22] and improved sensing in locations away from the muscle belly [4]. Earlier approaches measured skin deformation [15, 20, 27, 60, 70], but our band is tuned for fast-changing and highly-localized sub-millimeter wrist changes associated with grasping and interacting with objects, rather than global freehand gesture classification or on-body tapping. GestureWrist [60] presented 6 capacitive electrodes around the wrist to detect simple static poses; we go further with more receivers and more flexible detection. Photo-resistors have been used to also capture wrist contours, but are hard to scale; Fukui et al. [20] and ThumbSlide [2] used 75 and 16 photo resistors respectively and focused exclusively on capturing the index finger sliding along the thumb. EMPress [53] combined 4 force sensitive resistive (FSR) sensors with 4 EMG electrodes, and Liang et al. [44] used 5 capacitive pressure sensors. Liu, et. al. [47] explored accelerometer-based MMG to classify discrete gestures with a fixed wrist. While pressure- and inertial-based approaches use fewer sensors, they are still limited to a small set of gestures inadequate for capturing complex interactions with objects. Again, our work extends these...
earlier mechanomyography explorations by sensing force through tension, examining continuous range inputs in a ubiquitous environment, and using capacitance-based skin surface sensing easily scaled to high-detail wrist captures.

Some approaches have explored skin deformation and MMG principles at the back of the hand. Lin et al. [45] used four strain gauge sensors directly on top of the skin to detect hand gestures. Similarly, photo reflective sensor arrays on the back of the hand can measure detailed skin deformation for hand pose reconstruction [39, 68]; however these solutions impact hand mobility. More recently, vision-based systems with micro RGB [78, 81] or thermal cameras [31], mounted on the wrist facing the back of the hand, have been paired with CNN-based models for hand reconstruction. While these enable freehand interactions, leverage wide lenses embedded in a watch [78], and can be tuned for complex static poses such as sign language [10], they are still sensitive to occlusion from object interactions or finger-crossing and cannot measure internal state such as muscle tension.

2.2 Interaction with Objects in Ubiquitous Environments

Ubiquitous computing [75] proposes a rich set of interaction paradigms that move away from earlier "natural" computing ideas [3] in which wall-based computers reacted to pointing and voice commands [59, 72]. This newer vision suggests leveraging the affordances [21] of physical devices in UI design [65] and looking at everyday objects opportunistically as input devices [28]. Similar to Instant Controls (ICon) [12], Instant User Interfaces [13], and Ephemerous Interactions [74], we appropriate a wide range of everyday objects for interaction in our exemplar end-to-end system, but we avoid the need to physically augment the objects or environment. Object augmentation has been explored with visual markers [12, 28] and RFID tags [19]; alternatively, objects can be registered and tracked in real time using cameras [13, 74], but this limits the sets of objects to those registered with the system and requires an instrumented environment.

Mobile devices can overcome this limitation, with their multiple portable sensors enabling interactive everyday objects to be used peripherally, for example in an office desk or kitchen tabletop environment [57]. Instrumenting the fingertip with a micro-camera [79] or an RFID reader [71] has also been explored. Smartwatches can enable rich multi-modal inputs using microphones—such as slapping and banging the forearm, blowing, or tapping the foot [29]—in situations where one or both hands are busy.

Movement in space around the body can enable proprioception-enhanced inputs [11], further extending body-centric approaches [73] including for running or cycling [25]. Subtler interactions, like using the principles of magic to disguise embedded devices in everyday objects [1], are also possible, but again require instrumentation of the objects. While true micro-gesture— and grasp-based input has been examined through elicitation studies [5, 49, 64], it has been under-explored in implementation or achieved with limited setups like FingerInput’s use of an overlooking depth camera [66]. Our approach removes those constraints by instrumenting the wrist instead of the object and focusing on the dynamic use of everyday objects. This is also related to Affordance++ [48]—but with a focus on objects as input rather than output from objects—and further extends the idea of hands as a controller [5].

Previous work has often used surface mapping to create UIs on surfaces [26, 30, 77, 84]. We argue that this goes against the ubiquitous computing vision: these created overlays are similar to today’s mobile devices, centralizing interaction on augmented surfaces. Annexing Reality [30] and Gripmarks [84] are closer to our presented end-to-end system. Annexing Reality re-targets proxy shapes in the environment for haptic feedback in VR, and Gripmarks focuses on passive-but-graspable objects. They both appropriate surfaces for interaction and focus on shape affordance, while our approach leverages objects’ non-shape properties and interactive capabilities to enable ubiquitous inputs.

3 SENSING SKIN SURFACE DEFORMATIONS AT THE WRIST

3.1 Capacitive Sensing at the Skin Surface

Our proposed sensing technique uses parallel plate capacitors\(^1\): capacitance \(C \propto \frac{1}{d^2}\) the distance between the plates. We use the skin itself as a “plate”: movement and deformation of the skin causes a change in \(d\) (and thus \(C\)). Unlike accelerometers used in MMG [34, 38], capacitors can measure changes in \(d\) (e.g., vibration) without directly contacting the skin or dampening its motion. Laser-based range sensors can also measure displacement contact-free, but from a form factor perspective capacitive sensors are more compact than lasers.

3.2 Sensing At the Wrist

The wrist is our sensing location of choice: wrist-worn devices are unobtrusive and do not restrict hand motion or grasp. Additionally, watches are common wearable devices, which suggests that a sensing technique in this form factor has a path to wider adoption. With this location and form factor in mind, we discuss the input features that can be sensed at the wrist and relate them to underlying anatomical features.

The hand and fingers are controlled by tendons that pass along the anterior and posterior of the wrist to terminate with muscles in the forearm. The muscles and tendons of the upper limb typically exist in oppositional pairs, with flexors on the anterior of the hand, wrist, and forearm; and extensors on the posterior side. Abduction and adduction of the fingers is controlled by muscles intrinsic to the hand, i.e., that do not have a presence in the wrist. All thumb control tendons (flexion, extension, abduction, and adduction for all joints) pass through the wrist. The wrist proximal to the head of the ulna is home to the musculotendinous junction of these tendon/muscle pairs—the transition zone that connects pure muscle and pure tendon (see Figure 2).

When a muscle contracts and pulls a tendon, the musculotendinous junction moves, and the skin and subcutaneous fat layers nearby are pulled down into the space evacuated by the muscle-tendon unit. As this pliable potential space is filled or emptied, a user’s skin surface topography changes. Localized uplift or subsidence of the epidermis corresponds to specific tendons. For example, the extensor pollicis brevis and –longus are responsible for thumb

\(^1\)https://en.wikipedia.org/wiki/Capacitor#Parallel-plate_capacitor
The wrist posterior.

Wristband prototype uses Tactual Labs’ physical location of 16-bit capacitive magnitude signals, sampled while the topographical profiles study reported in Section 3.5 uses a 15x4 matrix of 4.75 mm x 4.75 mm electrodes (wider overall). These reported dimensions are the design used in our final studies, 4.2.1 use a 3x11 matrix of 6 mm x 6 mm electrodes (less density), which provides high frame rate, low latency measurements of capacitive signal strength based on the work of Leigh, et al. [43].

3.3 Wristband Prototype

Our wristband prototype, as noted, treats the skin as a plate in the parallel plate capacitor model. A high-frequency (254 kHz) electrical signal is “infused” on the skin by a soft pad covered in conductive fabric on the anterior wrist, and received by a 4x14 matrix of printed, flexible electrodes (each 4.75 mm x 4.75 mm, pitch 5 mm) statically positioned ≈2 mm above the skin of the posterior wrist (see Figure 3). These reported dimensions are the design used in our final studies, namely the Interaction Study in Section 4.2.2 and the Application Study in Section 5. Our grasp and force studies in Sections 4.1.1 and 4.2.1 use a 3x11 matrix of 6 mm x 6 mm electrodes (less density), while the topographical profiles study reported in Section 3.5 uses a 15x4 matrix of 4.75 mm x 4.75 mm electrodes (wider overall). These changes represent our evolving understanding of device signals; we consider the theory of operation and the connection between skin-surface measurement and the environment as more important contributions than a particular wristband incarnation. The wristband prototype uses Tactual Labs’ sensing platform, which provides high frame rate, low latency measurements of capacitive signal strength based on the work of Leigh, et al. [43].

Our studies focus mainly on extensor tendons, where the printed receivers are located. Flexor tendon motion is captured at a gross level: flexor motion changes wrist shape and coupling of the infusion pad, seen as a change in overall transmitted signal magnitude not localized to a particular matrix sensor. Inverting the prototype can swap the resolutions, but in conducting this work we observed the anterior skin has a larger range during interaction and more inter-user variation.

3.4 Signals for Sensing

Our prototype generates a 2-dimensional “heatmap” indicating physical location of 16-bit capacitive magnitude signals, sampled at 794 Hz. The band streamed data to the PC via Tactual Labs’ USB/Bluetooth drivers, which was processed using Python scripts.

Our data collection software, written in C++, allows defining a gesture list and recording duration. When starting a recording, a user is presented with a description and/or video of the gesture to perform; clicking a button begins recording. During recording, a live but abstract visualization of captured heatmap data is displayed. The tool continuously receives band information, stores it in a buffer, and writes to a timestamped CSV when recording is done.

3.5 Sensor Placement on the Wrist

We conducted a pilot data collection to understand our sensor’s sensitivity to different users and placements.

3.5.1 Procedure. Users placed the extra-wide (15x4) wristband with the matrix over the anterior of their right wrist. The wrist was held supinated (i.e., palm-up) with fingers extended flat. Users removed and replaced the band five times in the same location. We recorded one frame of data in each placement. Collections lasted less than 5 minutes.

We also used a conductive rod and gantry to test our sensor; measuring the rod’s physical position relative to the sensor’s reported signal allowed us to fit a power function mapping sensor signal magnitudes to distances from the sensor in mm. We applied this to our recorded data for analysis and visualization.

3.5.2 Participants. From our organization we recruited 10 participants (2F, 8M). Wrist circumferences varied between 152 and 205 mm (Mdn=186.5 mm, IQR=16.75 mm).

3.5.3 Results. We transformed our raw data into topographical maps with the aforementioned power function (see Figure 4). A pairwise Procrustes analysis between users’ first placements yields disparities of .01-.44 (M=.10, SD=.09); this large disparity suggests that using skin-surface profiles necessitates per-user calibration.

https://www.tactuallabs.com/
Pair-wise Procrustes analysis within-users yields disparities of .00–.18 (M=.05, SD=.04), suggesting similarity between placements but that models with small amounts of data may be unstable (see Figure 5). Between-user disparity was moderately correlated to wrist circumference difference, r(43)=.36, p=.01. We thus focus on single placements on single users as a base case for our technique. Also, our wide sensor, particularly on large wrists, seems to cause significant sensor-wrist contact at the ulnar and radial edges (while our goal is 2mm hover): we narrow our sensor in later studies to focus on the tendons.

4 INFERRING HAND INTERACTIONS FROM WRIST PROFILES

We now turn our attention to changes in skin surface topography caused by mechanical displacement and vibration of muscle contractions, as well as uplift and subsidence of moving or tensing tendons. This is closely related to Mechanomyography (MMG). We discuss features of objects and the environment which can be inferred using skin surface topography: specifically, we look at how pose, motion, and tension can aid in sensing the shape of and user’s interaction with an object. We thus explore a) measuring forces exerted at the fingertip, b) discriminating different grasps and gestures, and c) identifying states within ranges of gestures, across 3 user studies. All were conducted remotely in 2020/2021 during lockdown, leading to adaptations:

- users administered their own data collections using a mouse, and thus interactions were performed unimanually.
- each user had a personal, bespoke wristband, leading to signal and performance variation.

4.1 Tension → Force

Tendons and muscles are subjected to “loads” corresponding to weight or force, resulting in changed contraction profiles [67] or tension. Thus, musculotendinous activity is a proxy for force, which can be theoretically differentiated per finger (as finger tendons are independent). Load can be introduced on the muscles and tendons via either direct user actions (pressing on a table) or static object use anchored objects). This limited the grasps that we were able to test, as the variance in objects’ sizes to afford different grasps is large; we settled on 6 grasps using medium-sized objects.

Figure 6: Users in our force study used 9 springs with different maximum compression forces, plus a “no spring” 0 force case, which were all held with index finger tip and thumb tip .2” (≈5 mm) apart. Some users’ hands visibly deformed under different forces, as can be seen here with one participant in the 0 N (left), 22 N (centre), and 49 N (right) cases.

Participants: From our organization, we recruited 9 users (2F, 7M) whose ages range from 26–42 (Mdn=33, IQR=4) and whose wrist sizes range from 143–198 mm (Mdn=161.0 mm, SD=15.44 mm). All users were right-handed.

Results: Per-user compression force regression models had average error 13.66 N (SD=9.84 N) across all users.

Following Interferi’s weight regression and extrapolation analysis [33], we first performed tests in which, for each user, one set of 10 recordings was left out and regressed from the other four sets. This gave average error 12.39 N (SD=9.62 N) using scikit-learn’s MLPRegressor with default parameters, lbfgs solver, and tanh activation. We repeated the test, leaving out each force class and regressing it from the others using the same setup; this gave average error 13.66 N (SD=9.84 N). Although this error is relatively high, to our knowledge ours is the first method trying to derive pinch force with a hands-free wrist-based sensing approach, and we see value for discriminating more-tolerant force levels, e.g., button presses with different pressure intensities or rough assessment of strength. We expect these results to transfer to other grasps where users exert force involving other sets of fingers (e.g., squeezing a rubber ball or a sponge).

4.2 Pose and Motion → Shape and Interaction

Muscles and tendons in the wrist work together to move our fingers, and their pose correlates to hand pose, which is related to object shape. Their motion correlates to finger motion—whether user-initiated (I wiggle my finger) or environmentally-initiated (you wiggle my finger)—and to interaction patterns when an object or surface is involved.

We demonstrate mapping tendon and muscle pose and motion to user grasp, object shape, and interactions.

4.2.1 Grasp Study. While many grasp taxonomies exist [7, 14, 16, 46, 63], we focus on Feix, et al.’s [18]. We selected grasps covering power and precision, as well as differing opposition structures. To ensure we were sensing grasp rather than weight, we fabricated all objects from identical weights of identical materials and only tested objects that could be held entirely within the hand (i.e., we did not use anchored objects). This limited the grasps that we were able to test, as the variance in objects’ sizes to afford different grasps is large; we settled on 6 grasps using medium-sized objects.

Procedure: Using a uniform-density PlayDoh-like material, we created sets of four objects using molds and hand-shaping: a sphere (radius 4.25 cm), a cylinder (length 15 cm, radius 2.55 cm), a plate (length 13 cm, width 13 cm, thickness 1.8 cm), and a sphere with a
While LinearSVC was most accurate on average across all users, we recorded each grasp 10 times, for a total of 50 recordings. Each with default parameters, as our interest is in the data’s presence. Wrist sizes range from 143–198 mm (Mdn=161.0 mm, SD=15.44 mm). All users were right-handed.

Comparison to other wrist-based sensing techniques is difficult. Results: Per-user grasp classification had average accuracy 81% (SD=7%) across all users. Per-user held-object classification had a (predictably) higher average accuracy of 87% (SD=7%) across all users (see Figure 8).

We used generic machine learning models from scikit-learn with default parameters, as our interest is in the data’s presence rather than an optimally-designed model. We tested classification and regression trees (CART), linear discriminant analysis (LDA), logistic regression (LR), naïve Bayes (NB), linear support vector classification (LinearSVC), and support vector classification (SVC). For grasp classification, the best performers were LR (M=79%, SD=10%), LDA (M=80%, SD=11%), and LinearSVC (M=81%, SD=9%). While LinearSVC was most accurate on average across all users, for some it was not the most accurate model; allowing the system to use the most accurate model for each user bumps average accuracy to 85% (SD=9%) by mixing LinearSVC, SVC, LDA, and LR. For object classification, the results were, respectively, LinearSVC (M=84%, SD=7%), LR (M=86%, SD=6%), and LDA (M=87%, SD=7%). Comparison to other wrist-based sensing techniques is difficult as they usually discriminate dynamic gestures (which we investigate in the following section) instead of static grasps on objects, or do not control for weight when discriminating held objects. Most comparable might be the work of Fan, et al., [17] with EMG: they achieve 85% mean accuracy across 6 grasps similar to ours, although without controlling object weight as we did. For object recognition their weight-controlled study discriminating 4 sizes of 3 object types (spheres, cylinders, and plates) had 35% recognition accuracy; this is arguably a harder classification task than ours since their different-sized objects are grasped in very similar ways.

Model tuning and the weight/density/material variation of real-world objects would likely further improve our results, but they indicate distinguishing static grasps on objects with reasonable accuracy is possible. We changed our sensor after this study from a 3 x 11 matrix to a 4 x 14 matrix; more, smaller pads improve locational specificity in the data to capture subtler interactions.

4.2.2 Interaction Study. The more interesting corollary of “can we distinguish static grasps on objects?” is, of course, “can we distinguish dynamic interactions with objects?” We examined how pose, motion, and force can combine to represent a user’s interaction, and whether we could sense subtle changes in these dimensions from the wrist in a less-controlled scenario.

We focused on manipulative gestures (“whose intended purpose is to control some entity by applying a tight relationship between the actual movements of the gesturing hand/arm with the entity being manipulated” [58])—essentially the non-mediated version of “direct manipulation” [32] using real, physical objects. We looked explicitly at unimanual interactions for which the hand’s pose and motion (not location in space) uniquely determine state: these types of interactions can be sensed with a single wristband.

Manipulative gestures are a sort of subcase of grasp, but we turned from Feix’s grasp taxonomy to manipulation taxonomies, specifically Elliott and Connolly’s “Dynamic Hand Movements” [16] and Bullock, et al.,’s “Within Hand Prehensile Manipulation Taxonomy” [7]. Both categorize single-handed prehensile manipulations (i.e., hand/object interaction requiring more than one finger). We merge them to examine the following (see Table 1):

- **object contacts**: moving (hand touchpoints slide or rotate relative to the manipulated object) or fixed (touchpoints on the manipulated object are stable) [7, 16]
- **opposition** structure used: whether the force opposing the thumb is provided by the palm, finger pad(s), side of finger(s), or alternating contacts of side(s) and pad(s) [16]
- **motion type**: translation or rotation [7, 16]
- **axis** of the opposition vector in the hand coordinate system: X, Y, or Z [7, 16]

These dimensions can be extended from “describing the hand” to “using the hand as an input” by linking them with Card, et al.,’s classic input device taxonomy [8]. It organizes input device components by type and axis of movement (present in our manipulation taxonomy), and manipulation characteristics of the device; specifically whether using it causes positional or force change (a characteristic of the object being manipulated). Their final primitive is absolute...
Sensing Hand Interactions with Everyday Objects by Profiling Wrist Topography

Figure 9: Object interactions with their heatmap signatures. Images from the usability study of Ubiquitous Controls.

or relative measurement: normally a physical characteristic of an input device, we propose that in extending object manipulation to interaction it is a software mapping problem, which we explore in our application (Section 5).

Procedure: Users gathered objects from around their home with slide, rock, twiddle, squeeze, stretch and tripod pinch affordances (see Figure 9). For each interaction we suggested various objects. Users wore their wristband on their dominant hand to perform prompted transitions using their objects. Collections lasted 45–90 minutes.

We tested multiple states for each interaction: for all but rock we collected training data on 3 states (min, middle, max). Due to rock’s relatively larger range of motion, we collected 5 evenly-spaced states forming a superset of (min, middle, max). Users were invited to imagine a protractor or ruler aligned with their fingers to ensure state repeatability within a given interaction. Most objects were held entirely in the hand; for rock and twiddle the object (generally a jar or bottle) was set on the user’s desk.

Recording: Users recorded 2-second transitions between each pair of states for each range interaction. For the 5 interactions with 3 states we recorded each transition 3 times \( (\frac{5}{3})=6 \) transitions \( x 5 \) interactions \( x 3 \) repetitions = 90 recordings; for the 5-state range, we recorded one full set of transitions \( (\frac{5}{3})=20 \) recordings. This led to 110 total recordings per user. We used only the first and the last frame of the transitions as state samples for training and testing; this avoids overfitting consecutive frames, and we did not have continuous ground truth for intermediate states. For the 3-state ranges this yielded 36 samples (6 transitions \( x 3 \) repetitions \( x 2 \) frames), or 12 samples per state; for the 5-state range we got 40 samples (20 transitions \( x 1 \) repetition \( x 2 \) frames), or 8 samples per state.

All transitions for an interaction were recorded together, but object order and transition order within the interaction were randomized. Across all interactions participants leveraged 10 unique objects’ affordances, including reusing some objects for multiple affordances (see Table 1).

To assess if adding IMU information would improve classification results for dynamic interactions, we used an IMU 9DOF LSM9DS1 module with 3-axis accelerometer, 3-axis gyroscope, and 3-axis magnetometer in addition to the capacitive heatmap for this study. In contrast to the device—which samples at 794 Hz—the IMU is sampled at 952 Hz and multiple frames are batched if available. The IMU outputs both raw 16-bit data signals and pre-processed Euler angles, and we configure our accelerometer for \( \pm 2 g \) sensitivity for greatest distinctiveness in subtle motions.

Participants: We recruited 10 users (2F, 8M) from within our organization, ranging in age from 25–41 (Mdn=32, IQR=9.5), with wrist circumferences 143–198mm (M=165.5mm, SD=18.4mm).

Results: Using the same basic classifiers per-user described in the grasp study, we achieved 99 % average accuracy across users discriminating interaction types and 90 % classifying interaction state with LDA models. LDA achieves the best accuracies, which notionally makes sense as LDA assumes normal data distribution within and between classes—a natural fit with bodies and their motion. Training on IMU data as well improved interaction state classification accuracies only slightly (± 0.25 % on average per user and interaction). For interaction type discrimination, average accuracies were 99 % with or without IMU data. We hence report training results based solely on heatmap data in the following.

We evaluated the system’s discrimination ability for two classification steps (classify interaction, classify state within interaction). We performed cross validation on each of the 6 interactions (6-fold, step 1) split by its number of states (step 2). Given the characteristics of our collection, this lead to 8-fold cross-validation for rock and 12-fold cross-validation for each of the remaining 5 interactions. Samples per state were evenly stratified, with 1 in each fold. Thus we tested 8 classification tasks: 1 to discriminate the 6 interaction types, and 7 to discriminate range states within individual interactions (6 3-state interactions + 1 5-state rock).

We trained on just 10 samples per interaction type, selected in a stratified way across all states from the available sets of 36 or 40 samples. Our high mean accuracies of 99 % (SD=1 %) within users are similar to accuracies of other techniques inferring dynamic gestures from biometric sensing at the wrist, although this comparison needs to be considered with care as others use different gesture sets and do not include object manipulation. For example, Tomo [83] achieves 97 % accuracy on 10 hand gestures, Interferi [33] 93 % on 11 gestures, and Maereg et al.’s infrared-based approach [51] 98 % on 12 gestures, all within-users.

Models to identify state in the 5 3-state interactions trained on 10/12 collected samples per state, while for the two versions of rock we had fewer recordings available: these within-interaction state classifiers trained on just 6/8 collected samples per state. While 3-state gestures performed uniformly well across users with accuracies from 82–93 %, rock really excelled, with average classification accuracy 97 % (see Table 2). Rock has a larger interaction range and includes wrist rotation—which registers strongly on our prototype due to broad motion of underlying anatomy—while in other interactions the wrist is more stable.

To assess if our technique performs differently for people with a higher BMI due to fat layers on the wrist, we correlated recognition accuracies and wrist size (a proxy for user BMI). Each of our studies had users with wrist circumference >195 mm but we saw low correlation between size and accuracy \( (r^2 \text{ values: .04 (force), .02 (grasp), .02 (shape), .02 (interaction))} \).

5 UBIQUITOUS CONTROLS–LEVERAGING AFFORDANCES OF EVERYDAY OBJECTS TO CONTROL VIRTUAL RANGE INPUTS

We see two major opportunities to apply understanding of object properties and interactions with objects: 1) implicit activity recognition to infer user behavior and everyday activities like “drinking water from a glass” and 2) explicit gesture input to precisely control a virtual system.

In this paper, we focus on the latter. We present Ubiquitous Controls, a unifying interactive system based on the studies in Section 4,
which leverages the well-understood affordances of unaugmented everyday objects to provide guidance and haptic feedback for precise control of interactions. Many manipulative gestures are physically constrained to 1-dimensional motions within two extremes defined by an object’s affordance (e.g., opening a pair of scissors, or sliding a box cutter), which makes them map well to virtual 1-dimensional inputs, like sliders or dials.

The system works by first differentiating between a variety of dynamic hand interactions with objects and then regressing the position within the range of the detected interaction. We developed a machine learning pipeline to run alongside the dedicated data collection software described in Section 3.4.

We now provide an interaction scenario, system overview, and evaluation with 8 participants controlling 3 widgets.

### 5.1 Using Ubiquitous Controls

A user reading a book on their couch wants to turn their music down. Instead of standing or searching for their phone, they perform a distinctive gesture to activate the Ubiquitous Controls system and twiddle the lid of their water bottle, at hand, counterclockwise to reach the desired volume (see Figure 10).

For the interaction to work, the user needs to configure their water bottle to be an input device; this task occurs only once per object, on-demand or ahead of time. They collect samples of their hand twiddling the cap to its extremes (i.e., far left, far right, centre); this data collection takes approximately 1 minute. The data are sent to a classification optimizer, which compares several machine learning configurations to select two models. One separates “twiddling a bottle cap” from other recorded controls, like “squeezing a toy” or “sliding a hair dryer selector.” The second regresses the position of the bottle cap interaction based on the three recorded states.

After model training and selection, the system awaits the interaction-triggering gesture by analyzing live IMU information. Once detected, wrist profile and IMU data are sent to the first classifier to determine interaction, and to the interaction-specific model to regress range position. Finally, output is sent to the appropriate application: in this case, it adjusts the music player’s volume level.

### 5.2 Software Implementation

Based on the results of our Interaction Study, for range position regression we collect 1-second recordings of holds in three positions: min, middle, and max.

Our training script collects the output csvs and extracts features. Cross-fold validation, per user, determines the best model. The live script takes this model and runs live data through it, outputting regression or classification information. Both scripts are written in Python with scikit-learn [56]. Our two model stages (1) classify high-level interactions and (2) regress a particular range location.

We use raw heatmap, accelerometer, and gyroscope values combined with 23 statistical features including z-score, mean, and standard deviation. As in the study, we compare LDA, CART, LR, NB, and LinearSVC classifiers for gesture classification and use a linear regression model for range interpolation. We use standard configurations for all models. Trained models are n-fold cross-validated (where n=number of samples/number of trained states): the best classifier and the linear regression model are stored for live use.

In live mode, we detect a “double-flip” gesture, a placeholder activation trigger originally designed as a robust input-delimiter for mobile motion-based interactions [61], by thresholding the IMU gyro magnitude; the timeout for completing both flips is 300ms. Our system then constantly predicts interaction type and position until detecting the next “double-flip” gesture, which deactivates the interaction mode. The stored classifier determines the current interaction using single heatmap+IMU frames (see Figure 9 for example signatures).

We smooth regressed range position with an empirically-tuned 1-Euro Filter [9]. The live classification/regression script samples data from the band at ≈60 Hz and makes constant predictions in real time: we have achieved this on a variety of desktop and laptop computers, including a 2014 i7-4790 processor with 32GB of memory.
Extracted data are sent to applications as input. We purpose-built several widgets for Ubiquitous Controls, but with Sikuli [80] or other end-user programming techniques it would be possible to integrate such inputs into more applications.

5.3 Evaluation

We assessed the usability of our Ubiquitous Controls system in a real-time scenario with a study in which 8 users interacted with three widgets. Our primary goal was testing a live instance of our pipeline and whether it was functional to enable ubiquitous, continuous control input using affordances of everyday objects. Our hypothesis was that:

Hypothesis 1. The Ubiquitous Controls pipeline enables users to input interpolated range values within 1.5s

This is intended to mirror prior findings that users can take 1.172 s–1.233 s from stimulus to target acquisition using a trackpad with their dominant and non-dominant hand [36], or that a typist using an unfamiliar keyboard can require 1.2 s per key entry [50]. Participants completed input tasks in median 1.2 s per prompt, providing evidence for Hypothesis 1.

Asked about their preferred interaction technique, 3 users mentioned Ubiquitous Controls, 3 preferred freehand gestures, and 2 declared it depends upon the situation. Users reported state reproductibility and haptic feedback as favorable aspects of object-supported interactions, suggesting our proposed interaction technique is a good alternative to freehand gestures for some kinds of input.

5.3.1 Procedure. We asked users to gather interactive objects from their homes; we provided them with a recommendation list illustrating different affordances. In contrast to the Interaction Study, we did not require specific affordances: users were explicitly allowed to use any objects they wanted as input controls. Studies lasted up to 60 minutes, exploring 5 tasks to be performed using both interaction methods (object-supported interactions and freehand gestures). Each participant wore a different wristband of the kind described in Section 3.3. A participant wore the same wristband for both input gesture techniques. The task consisted of controlling 3 widgets: a horizontal slider to "control the seek position of a video," a zoomable canvas to "control the zoom level of a map," and a knob to "control the volume of a stereo" (see Figure 11). Before each task, the experimenter described the widget and required users to define a freehand or object-supported gesture to control it, with users eventually doing both for all 3 widgets. The overall widget x interaction method configurations were randomized.

Every task consisted of five phases: a definition of the gesture’s 3 distinct states (max, min, and middle), a data collection, an instantaneous training, an experimentation phase where users could try out their trained model on the widget, and a test phase where users input prompted values as quickly as possible.

Based on the findings of the Interaction Study that 10 samples per state provided satisfying classification accuracies to discriminate 6 gestures and 3 states within each gesture, we had users collect 30 training samples per task (3 states x 10 repetitions). The collection comprised 1s-long static poses rather than transitions, and lasted roughly one minute for each 30-sample set. The user then started a script to train a model with these samples, forcing them to put down their object or change their hand posture prior to the experiment and test phases. The training script generated a linear regression model, which interpolated values within the range bounded by max and min; this permitted a richer, more continuous interaction compared to the 3-state classification models from the Interaction Study. The output of the regressor was not smoothed or post-processed, but was used raw to map inputs to the virtual control range. Prior to the recorded test phase, the user experimented with their trained model and its mapping to the 2D graphical widget for a maximum of two minutes. During this phase, if a user was not able to perform controlled manipulations (i.e., all inputs were regr essed as "out of range," either above the trained maximum or below the trained minimum), we allowed them to recollect data and train again. The test phase consisted of a given number of virtual control value prompts to be matched within two minutes. Each prompt appeared the same number of times in a randomly-shuffled sequence. Holding the correct value for a fixed period of time triggered a transition between prompts. For the knob and zoomable canvas, users held 25 prompts for .5s each; with visual prompts and regressed model output both snapping to 1/5th of the overall input range. For the horizontal slider, users held 50 different prompts for .5s each, where each prompt was represented by an interval of 1/10th of the overall input range. Regressed values were presented continuously on the UI (see Figure 11). Performance metrics and sensed wristband data were captured during the study.

5.3.2 Participants. We recruited 8 users (1F, 7M) within our institution. Their ages ranged from 24–45 (Mdn =33.5, IQR=5.75) and their wrist circumferences from 152–210mm (Mdn=170.5mm, SD=19.4mm).
Figure 11: We used 3 UI widgets in the usability study to assess performance of our system: a horizontal slider (A), a zoomable canvas (B), and a rotary knob (C). For the slider the prompts were given in intervals and the output of the regressor was continuous (not snapped). The canvas and the knob featured visual “snapping” of both prompts and inputs to 5 evenly-spaced locations.

Figure 12: Users in our Ubiquitous Controls evaluation had various interesting interaction strategies. One used an unpowered oscilloscope (A). Another used both hands to precision-set a combination lock (B). Another precision input was a non-digital pair of brass calipers (C). One user’s “squeeze” object, a thin-walled plastic cup, broke through use as an input (D).

All but one of the studies was performed remotely, with the in-person study conducted in a physically-distanced manner outdoors.

5.3.3 Design. We recorded 23 trials for object-supported interactions (3 widgets x 7 users + 2 widgets x 1 user) and 22 trials for freehand gestures (3 widgets x 6 users + 2 widgets x 2 users). Due to timing and technical issues, 2 users could not complete the full study; one completed 2 widgets per interaction method and the other did not perform the final freehand interaction. In total, through 45 trials our users leveraged 13 unique object types, all 6 unique object interactions that we tested in our Interaction Study, and 8 unique freehand gestures for their controls. Of those 45 trials, the data collection and model training phases were repeated for 4 due to poor system performance during user experimentation. We define one complete test phase trial as 25 completed prompts within 120 seconds; this truncated the horizontal slider conditions from 50 prompts to balance our analysis.

5.3.4 Results.

Quantitative. With 407 matched prompts across all three widgets, most users could successfully control all widgets using object-supported interactions. For 3 users input for the range slider with 10 buckets did not work during the study; without regression smoothing the input jitter made a .5 s hold challenging. Users matched average 21.4 / 25 prompts (Mdn=25, SD=6.1) within the time limit, for interactions with at least one matched prompt. For such trials all but 2 per widget were finished in time.

We found strong support for Hypothesis 1: median input time with Ubiquitous Controls was 1.2s, with 58 % of prompts completed within 1.5 s (not including the required .5 s hold). As with other devices, more practice with the inputs and their behaviour would likely reduce input time. Though the slider had twice as many buckets as other widgets, no interaction was significantly slower or faster across trials with ≥1 matched prompt.

Mean classification accuracy for the 3 range states was 92 % (Mdn=97 %, SD=9 %), similar to the Interaction Study. We also included IMU data alongside heatmap data for training and testing. Similar to the Interaction Study, removing IMU snapshots reduces performance only slightly (-0.41 % on average per user in post hoc analysis). Mean range interpolation error was 16 % (Mdn=10 %, SD=14 %), similar to Interferi [33]. This error also represents user interpretation error: users estimated state positions during data collection and we had no ground truth. For interaction discrimination, the mean accuracy of 84 % (84 % without IMU, Mdn=99 %, SD=28 %) was lower than in the Interaction Study; as some participants used the same object-supported interactions for different interaction tasks, we believe the median of 99 % is a more appropriate metric.

Users were also able to control the widgets using freehand gestures; there were no significant differences between object-supported and freehand gestures in input time, number of matched prompts, model fit, or interaction type discrimination accuracy.

Qualitative. Participants’ preferences about the interaction methods were mixed: three users preferred object-supported interactions, three preferred freehand gestures, and two declared it depends on the situation. Many users highlighted state reproducibility and haptic feedback, saying the “force made it easier to control” (P7) pinching with a rubber band versus freehand, or that feedback from a screw-top lid “lets you know where you are on the thing” (P4).

Many users explicitly mapped an object interaction or gesture to a widget’s visual appearance or existing use in the wild. Users mentioned “if it’s a video player, I’d like it to be linear” (P4), or “I’ve turned the knob of a stereo, so I know that a bottle cap is more consistent with the experience [than a linear control]” (P3).

Our system’s treatment of object-supported interactions did always not map well to users’ understanding of how inputs work. For example, P1 used a knob on an unpowered oscilloscope as an input: in training, they twisted using mainly their wrist, and in testing with mainly their fingers. In normal use, these methods behave identically, but in our system they do not (see Figure 12A).

On using objects as inputs, one user suggested they would carry a “collection of favourite objects that were comfortable to use,” or a multi-function input object (e.g., fidget cube) designed for this type of interaction (P6). Conversely, another mentioned that any object they have has to be “worthy of carrying with me,” suggesting that an object only for input would not be worth the space in their bag or pockets (P2). In the end, it seems both object-supported and freehand gestures have a role in pervasive interaction, but what defines their roles requires further study.

6 DISCUSSION AND LIMITATIONS

We have demonstrated inference of object properties using capacitive topographical changes at the wrist, including an end-to-end system leveraging them for explicit ubiquitous input. As demonstrated through our studies on force, grasp, and interaction, tapping into our body’s own actuation systems to detect isometric tension, pose, and movement re-imagines the hand as a sensor instead of integrating electronics into every object we wish to imbue with
interaction. This also leads to the challenge of anatomical differences between users, which we showed extends beyond wrist circumference to wrist profiles. We now summarize challenges and opportunities that surfaced in executing this research.

6.1 Performance

While 3-state classification and 5-state interpolation with our prototype pipeline was usable, users’ experiences with 10-state interpolation were mixed, and 4.9% of successful object-mediated inputs with Ubiquitous Controls took over 10 seconds to complete. We anecdotally noted these issues were often due to regression instability. In the study, we did not filter regressed data: smoothing with an empirically-tuned 1-Euro Filter [9] has dramatically improved system usability. A post hoc analysis on recorded frames from the study reduced mean “input” times by .21 s, even accounting for lag introduced by smoothing, and 1 percentage point fewer trials took >10 s to complete. The filter reduced regressor noise, but further work is required to ensure robust, continuous, and precise input.

Our sensor design evolved throughout the work and we used an IMU for some experiments. Post hoc analysis revealed IMU contribution to accuracy was minor (+.25% in the Interaction Study, +.41% in the Application Study), and future work can explore new capacitive matrix geometries and densities to improve accuracy.

While our first study shows anatomical differences between people, we believe it possible to build a population-level machine-learning filter to separate arm/wrist motion and position from finger motion and position in order to better match users’ mental models of how Ubiquitous Controls inputs work and to make interactions transfer across users. This would require significant data collections on people of varying wrist characteristics and is outside the scope of the present contribution.

6.2 Midas Touch

As computation moves to the body and everyday objects may serve both analog and digital functions, we encounter the so-called “Midas Touch” problem, where a system must distinguish between functional interaction and intentional input with an object [13]. In our prototype system, we use double-flip to demarcate these modes, but further exploration may find contextual or physiological (e.g., EEG) signals to distinguish whether a user twisting a bottle lid intends to take a drink or control their podcast. The ideal solution does not require a specific initiation gesture, but this is dependent upon the available cues. Multi-modal techniques like voice or eye-tracking might be required to disambiguate functional interaction and intentional input. Without extra modalities, a system could track events before and after recognition to disambiguate usage, however this introduces input delays as the system waits for differentiating events after recognizing an action.

6.3 Beyond Shape and Force

Better understanding the objects a user is interacting with is an obvious future direction that builds on the primitives of shape, force, and gesture. Force combined with pose and motion may reveal surface hardness, springiness, or mechanical stickiness. Touch on a held object from another hand may also be visible [82]. Taking motion’s derivative, acceleration profiles during object interaction may correlate to weight, centre of mass, or change in weight distribution (e.g., if the object contains liquid).

6.4 Reuse and Assemblage of Other Devices

A corollary of re-purposing existing object properties is reprogramming existing input devices for new uses. After all, a disconnected game controller is just a high-quality fidget cube: both are primed with interesting physical mappings. A user could map the brightness of their room to the roll of a mouse wheel, adjust which cooktop burner is active by thumbing a joystick, or reuse a favourite PS/2 peripheral with a new machine without an adapter.

For accessibility reasons, allowing a user to reuse an object or input device they are familiar with for new and arbitrary interactions is a boon. Relocating a component’s sensing from the device to the user also enables a new way to prototype input devices: assemblage of existing affordances held together by clay or tape is easy to reconfigure and immediately test, similar to the concept of Makers’ Marks [62].

6.5 Dynamic and Configurable Mapping

The distributed nature of input and output also introduces challenges of mapping, for which we need new paradigms in end user programming that allow tying a specific input to a specific output.

All our tested interactions include explicitly-coded mappings, with the minimum value of a physical range permanently connected to the minimum value of a digital range, and all interactions are 1:1. More subtlety may be desirable: the abilities to map sub-ranges, to dynamically map physical and digital inputs, or to choose a different gesture according to the situation (working out vs. lying in bed) each contribute to a personal mapping experience where the interface bends to meet the user’s intentions and is not bounded by virtual inputs or specialized input devices.

Meaningful, pervasive mapping techniques with Ubiquitous Controls also open opportunities for more hygienic and personal interactions with public displays or infrastructure: one’s own items can be leveraged in lieu of a public touchscreen or other input device.

7 CONCLUSION

We have proposed a capacitance-based sensing method to extract the surface profiles of anatomical structures in a user’s wrist; further we described our prototype implementation of a wristband using this sensing technique. We discussed the implications of the actual anatomy being sensed: that tendon and muscle tension, pose, and motion in the location of our wristband can be mapped to, respectively, force, grasp, and interaction of the hand and fingers with external objects. We demonstrated these mappings through several atomic studies. Finally, we describe Ubiquitous Controls, an end-to-end system that uses these capabilities to let users control virtual range inputs by interacting with unaugmented objects, and we conclude with a discussion of the implications and opportunities of using the hand as a sensor. We anticipate that using anatomy to understand interactions with and properties of everyday objects will provide a powerful framework for ubiquitous computing.
ACKNOWLEDGMENTS
We would like to thank Maria Ruiz Maya, Goutham Palaniappan, and Kevin Stager for their assistance in preparing (and repairing) hardware prototypes; Myles Gifford, Pablo Saldarriaga, Michael Mohan, and Peter Hamilton for their support of the software tools, all our anonymous participants, and Cliff Forlines for his guidance and shepherding of the project.

REFERENCES


