Headar: Sensing Head Gestures for Confirmation Dialogs on Smartwatches with Wearable Millimeter-Wave Radar

XIAOYING YANG, University of California, Los Angeles
XUE WANG, University of California, Los Angeles
GAOFENG DONG, University of California, Los Angeles
ZIHAN YAN, Massachusetts Institute of Technology
MANI SRIVASTAVA, University of California, Los Angeles
EIJI HAYASHI, ATAP, Google
YANG ZHANG, University of California, Los Angeles

Nod and shake of one's head are intuitive and universal gestures in communication. As smartwatches become increasingly intelligent through advances in user activity sensing technologies, many use scenarios of smartwatches demand quick responses from users in confirmation dialogs, to accept or dismiss proposed actions. Such proposed actions include making emergency calls, taking service recommendations, and starting or stopping exercise timers. Head gestures in these scenarios could be preferable to touch interactions for being hands-free and easy to perform. We propose *Headar* to recognize these gestures on smartwatches using wearable millimeter wave sensing. We first surveyed head gestures to understand how they are performed in conversational settings. We then investigated positions and orientations to which users raise their smartwatches. Insights from these studies guided the implementation of *Headar*. Additionally, we conducted modeling and simulation to verify our sensing principle. We developed a real-time sensing and inference pipeline using contemporary deep learning techniques, and proved the feasibility of our proposed approach with a user study (n=15) and a live test (n=8). Our evaluation yielded an average accuracy of 84.0% in the user study across 9 classes including *nod* and *shake* as well as seven *other* signals – still, speech, touch interaction, and four non-gestural head motions (i.e., head up, left, right, and down). Furthermore, we obtained an accuracy of 72.6% in the live test which reveals rich insights into the performance of our approach in various realistic conditions.

CCS Concepts: • **Human-centered computing → Gestural input**; **Interactive systems and tools**.

Additional Key Words and Phrases: Wearable interaction, Gestural input, Millimeter-wave radar, Head gestures, Smartwatch

1 INTRODUCTION

Smartwatches are becoming increasingly intelligent – thanks to the rich user activity information acquired by learning-based inference of various smartwatch sensors (e.g., IMUs [42, 44, 58], GPS [45], microphones [44, 73], PPG sensor[13]), many applications on smartwatches have retired complex interaction sequences by proactively making suggestions and taking decisions from users. From approve/deny service recommendations to start/end

Fig. 1. Examples of confirmation dialogs on smartwatches. First row from left to right: accept/dismiss a phone call, resume/end navigation, snooze/stop an alarm, start/stop an alarm, archive/dismiss an email, start/stop an emergency call, reply/dismiss a message. Second row from left to right: delete/keep an app, confirm/dismiss a reminder, confirm/dismiss a note, approve/deny two-factor authentication, turn on/off a setting, accept/cancel a service change, end/pause a workout timer.

exercise timers, intelligence on smartwatches could overcome the limited interaction bandwidth constrained by small screens, by narrowing down users' input options into only two categories – *accept* or *dismiss* the recommended actions. We refer to interfaces that ask users for accept/dismiss responses as *confirmation dialogs* in the rest of this paper.

In fact, such confirmation dialogs have been used in a wide range of scenarios. For example, a smartwatch user would use confirmation dialogs to accept/dismiss a phone call, approve/deny an authentication request, stop/snooze an alarm and beyond, as shown in Figure 1. Prior work has studied the efficacy of confirmation dialogs and demonstrated they can reduce user errors [61] and have been widely practiced by smartwatch developer communities [7, 8, 70].

To facilitate natural and efficient interaction with confirmation dialogs on smartwatches, this work turned to head gestures for solutions. We aim to use the two natural head gestures – *nod* and *shake*, to replace clumsy touchscreen interactions which is the status-quo interaction method for confirmation dialogs, allowing for swift interaction sequences that could make smartwatches more intuitive to use in a wide range of use scenarios. We built our head gesture detection technique off of millimeter wave (mmWave) radars which yield high-fidelity spatial information without revealing unintended sensitive information in the background, mitigating user concerns about privacy. More importantly, mmWave radars can be compact and integrated into smartwatch bezels [1] or potentially fabricated as smartwatch straps [83]. The feasibility of this sensor on devices with small form factors has been demonstrated by Google Project Soli [50, 78], Google Nest, and Google Homes. Building upon existing efforts of using mmWave radars to sense gestures for explicit input in HCI [35, 50, 78], this research investigated a new gesture set – head gestures, on a wearable platform, which demands new signal processing, inference pipeline, and evaluations. This is uniquely challenging due to the hard-to-assume relative positions between a user's head and the radar sensor since users hold smartwatches differently, with use scenarios also impacting relative placement. What makes it even more challenging is the constant motion of the user's hand hovering in the air, causing adversarial noise with similar characteristics as head gestures. In addition, the variation

---

[1]Google's smartwatch with radar for gesture control https://youtu.be/jWNebDDmuXc

among users performing head gestures in terms of duration, magnitude, and velocity makes generalization of the system a real challenge. These unknowns and challenges highlight our contribution.

On the other hand, this work fulfilled a missed opportunity of leveraging the intuitiveness and universality of head gestures to enrich the interactivity of smartwatches. It is intriguing to find the name *"dialog"* commonly appears in interfaces on smartwatches, contradicting the perceptual capabilities that they rely on – contemporary smartwatches predominantly rely on touch interactions. This touch-based interaction modality precludes the opportunity for smartwatches to understand their users like how people understand each other in conversations. On this note, our research aims to allow smartwatches to better understand their users by recognizing natural head gestures which virtually all smartwatch users know how to perform. Anecdotally, people start performing head gestures long before their first conversations in social settings. Charles Darwin's observation documented that child shakes their heads in order to stop food intake when sated in breastfeeding, and thus establishing a connection between rejection and the shake gesture [15]. We aim to enable such natural means of communication between users and their smartwatches.

Despite the merits of head gestures, few prior systems have shown their detection on wearable platforms. Closest to our research is the work by Raja et al. [65], which is an RF sensing system for dynamic head gesture recognition using Doppler shift (i.e., relative velocity of head to the radar). Most studies were done with the radar transmitter and receiver detached and located at different locations in space, a possible configuration for deployed sensors (e.g., smart environment) but unfitted for wearable scenarios which our research targets. In addition, including the aforementioned related work, no prior research has investigated the configuration of the smartwatch and its user's head such as their relative position, motion in the wrist-raise-activation sequence, to which our survey study uniquely contributes. We conducted a series of modeling and simulation to verify the sensing principle – RF responses per the relative head motions with respect to the radar sensor, as well as a survey study that investigated head gestures (i.e., periodicity, magnitude, speed) which yielded insights that allow our sensing pipeline to best detect *nod* and *shake* gestures close to ones performed in natural conversational settings. Finally, we created a custom signal processing and deep learning (DL) pipeline which we evaluated with a user study of 15 participants and a live test of 8 participants in various realistic conditions, yielding an accuracy of 84.0% and 72.6% respectively.

Overall, our observation is that head gestures are simple, natural, subtle, and hands-free interaction modalities that virtually every smartwatch users understand and practice in daily conversational settings. This research utilizes head gestures to enrich the interactivity of smartwatches with a practical solution using wearable mmWave sensing. Our contributions are as follows:

- Realization of head gesture recognition using wearable mmWave sensors.
- A series of modeling and simulation that verified RF responses induced by head motions.
- Two survey studies that investigated the variance of user head gestures, and relative smartwatch position and orientation with reference to a user's head.
- An end-to-end signal processing and deep learning pipeline.
- Two user evaluations that proved the feasibility and revealed rich insights into the performance of our technique in various realistic conditions.

2 RELATED WORK

2.1 Enrich Interactions for Smart Watches

Efficient approaches to expanding input vocabulary for smartwatches have long been sought after to overcome limited screen sizes and the resulting "fat finger" problem. One approach is to decrease the touch point size using stylus [84], while others seek to expand the input area [37, 43, 100, 105]. Camera-based approach, RF sensing and laser projectors have been adopted to transform user's skin into touch input surfaces in prior work such as

Skinput [34], SkinTrack [100] and Skin Buttons [43]. Built-in sensors on smartwatches including microphone and IMU have been repurposed to offer on-skin gesture recognition capabilities [95].

Prior work has also created interactions beyond touch to overcome the above challenge. For instance, WatchMI proposed pressure touch, twist, and pan gesture input to enhance touch interaction [92]. SoundTrak expanded human-wearable input operation space from two to three dimensions by tracking 3D trajectory of the finger on air [96]. In addition to these interaction techniques that would require both hands of a smartwatch user, one-handed techniques have also been explored. Joystick-like motions of the wrist have been used for controls input [29, 30] and text entry [28]. Finger tracking has been explored by capturing bio-acoustic signals of fingers to enable hand gesture detection such as pinch and flick as shown in ViBand [44]. Magnetic sensing techniques are also popular [17, 66]. For example, AuraRing [69] features a ring-based electromagnetic transmitter coil and wristband-based multiple sensor coils to track fine-grained finger movements with a 0.1 mm resolution. MagTouch [64] leveraged built-in magnetometers on smartwatches to identify and track a user's fingers instrumented with small magnets.

Compared with these prior systems which extend interactivity to surrounding regions around smartwatches, we extend it further to a user's head. Head-based interactions are simple to perform while freeing up one user hand [36]. Head-oriented signals such as voice commands [67, 71] and gaze [6, 22, 33, 86, 87] have been leveraged to realize hands-free interactions for wearable devices, but these techniques are drastically different than head gestures which we focus our sensing approach on.

2.2 Sensing User Head Gestures

Our heads, even as simple as their orientations, convey much information in daily communication. Much prior work has leveraged signals from users' heads, in particular facial gestures [19, 20, 48, 88, 99], for interactive systems. These head-instrumented wearables often hold a relatively fixed position with the head and sense nuanced signals centered around a user's face. We focus on prior work that senses user's head gestures next.

Nod to Auth [79] used IMU sensors to reconstruct trajectories of head gestures such as nod and shake of a VR/AR user for authentication. Pose-on-the-Go [4] and ControllerPose [5] took a hybrid approach using vision and inertial sensing to track human pose. Kapoor et al. [40] used an infrared camera deployed in the environment to track nod and shake by tracking pupil movements. A more common vision-based approach tracked head gestures by tracking the motions of a user's head as a rigid body [23, 60, 81].

Closest to our work is a system using mmWave radar for 3D head motion detection, presented by Raja et al. [65], Bresnahan et al. [14] and Chae et al. [16]. These work investigated a driver head motion monitoring system and experimented in a stationary setting. Foremost, the radar position with respect to the user head is constant and is known as prior knowledge. However, this assumption does not hold in the case of smartwatch uses, to investigate which, we conducted studies to survey users' habits of positioning and orienting their smartwatches. We also proposed and evaluated new inference techniques in dealing with the complexity of positions and orientations in smartwatch head gesture sensing.

2.3 mmWave Sensing in HCI

Closer to our work on the technical front are previous HCI systems that also leveraged mmWave sensing. In the commercial domain, this technology has mostly been used for for automobiles (e.g., collision prevention [104], cruise control [53]). In this section, we focus on existing systems that have HCI applications. For example, researcher have instrumented mmWave radars in environments to sense states of interactive controls [89], occupancy [46], and even user emotions [31]. This technology has also demonstrated potential in material detection. For instance, RadarCat [91] is able to identify materials and objects in real-time. FG-LiquID [49] detected the properties of liquid such as alcohol content, milk quality, and water contamination.

Additionally, mmWave radar has been a popular sensing solution in human activity recognition (HAR) for being insensitive to unintended information (i.e., RGB images of background) in user environments. For instance, RadHAR [74], a framework that used sparse and non-uniform point clouds, could perform accurate HAR. Vid2Doppler [3] synthesized mmWave radar data using videos of human activities for effectively training privacy-preserving HAR. IMU2Doppler [11] provided another way to deal with limited labeled data using IMU datasets acquired by off-the-shelf smartwatches. mID [103] tracked and identified multiple people simultaneously with a mmWave radar. Xie et al. [85] proposed a universal targeted attack method with the goal to protect mmWave-based HAR systems from adversarial attacks.

Additionally, mmWave radar has been used to detect gestures and poses at full-body scales. For instance, RadarNet [35] can effectively recognize directional swipe gestures (i.e., up, down, left, right) and omni-swipes (i.e., swipes in any direction) via a miniaturized 60 GHz radar sensor. Zhao et al. [101] realized through-wall human pose estimation leveraging the phenomenon that wireless signals with specific frequencies can travel through walls but reflect off of human bodies. Prior work has also used mmWave radar to track parts of a user's body such as arms [51, 63], forearms+hands [72], and fingers [80].

Closest to our technique are mmWave sensing systems with a wearable or mobile form factor. Johnson et al. [39] presented a mmWave system for accurate arterial pulse waveforms measurements without contacting the pulsing region. Besides, this sensing technique could enable gesture sensing even if the sensor is occluded by fabrics [47]. Finally, Google Soli [50] proposed an end-to-end system tailored to close-range micro hand gesture sensing. This system introduced seminal radar design principles, a hardware abstraction layer (HAL) and architecture, and gesture vocabularies for interaction, to support hand gestural interactions with mmWave sensing on wearable and mobile devices.

3 SENSING PRINCIPLE

In this section, we address the sensing principle our system leveraged for recognizing user head gestures. We started by describing the source of signal resulting from users' head motion and briefly discussing Frequency-Modulated Continuous Wave (FMCW) mechanism. To verify our understanding of the sensing principle, we conducted simulations by modeling a user's head and a radar, the result of which is compared with measurements from a real mmWave sensor.

3.1 Source of Signal and FMCW Sensing

3.1.1 RF Reflection off of Human Skin. Similar to other forms of electromagnetic waves (e.g., light), RF signals (20 kHz - 300 GHz) can undergo reflection, refraction, diffraction and scattering when encountering certain interfaces, depending on wavelength, incident angle and medium properties on both sides of the surface. Conductive materials (e.g. metal) have high reflectivity and reflect almost all incident waves, while dielectric materials permit wave penetration to a certain extent and thus yield larger refraction intensity. The conductivity of a material can be measured by $\sigma = 2\pi f \varepsilon_0 \varepsilon''$ in Siemens/meter (S/m), where $\varepsilon_0 = 8.85 \times 10^{-12} F/m$ is the vacuum permittivity, f is the operating frequency of radio waves and ε'' is the imaginary part of relative complex permittivity of a material given by $\varepsilon = \varepsilon' - j\varepsilon''$. The relative permittivity of human skin at millimeter wave frequencies is estimated by modeling and measurements in prior research such as [1, 25, 82, 94]. To model a user's head, we refer to [9] to obtain the relative complex permittivity of dry skin, specifically $\varepsilon = 6.57 - j8.91$ at 77 GHz.

Due to the short wavelength of millimeter waves compared with human body, at the air-skin interface, human skin is usually modeled as flat surfaces and can lead to reflection and refraction behaviors of incident waves. According to Fresnel equations[2], the power reflection coefficient and power transmission coefficient can be measured by R and $1 - R$, where R is given by Equation 1 and θ_i is the incident angle [82]. Overall, we calculate

---

[2]Fresnel equations https://en.wikipedia.org/wiki/Fresnel_equations

that at 77 GHz, skin surfaces reflect an approximate 58.34% of normal incident power, confirming the reflectivity of human body to RF signals at millimeter wave spectrum.

$$R = \left| \frac{-\varepsilon \cos \theta_i + \sqrt{\varepsilon - \sin^2 \theta_i}}{\varepsilon \cos \theta_i + \sqrt{\varepsilon - \sin^2 \theta_i}} \right|^2 \tag{1}$$

RF reflection is often modeled by a mix of reflective and diffusive reflections [21]. Reflective reflections are angle-dependent and have been leveraged to create reflectors that reflect RF waves to their sources for radar visibility (e.g., corner reflectors in boats for radar detection and navigation). Most everyday surfaces yield diffuse reflection and reflect RF waves in multiple directions. In this research, both reflective and diffusive reflections result in a variance of reflected RF energy in *Headar* when a user's head is angled differently during head gestures. Our goal was to leverage the difference between variances from *nod* and from *shake* to detect head gestures. However, we first conduct simulations to verify the existence of such a difference, which also verifies our understanding of the sensing principle.

3.1.2 FMCW Mechanism. Since much literature in HCI has documented FMCW radar sensing [26, 35, 50, 98], we will not detail its principle of operation. In short, FMCW radar offers high range resolution and allows real-time updating of measurement by continuously transmitting signals with varying frequency, making it more applicable to interactive devices. Also, FMCW radar eliminates the need for the high frequency sampling in measuring the time of flight (ToF) by converting ToF into frequency components, which could be calculated by Fast Fourier Transform (FFT) to estimate range and velocity. Beam-forming techniques with multiple input and multiple output (MIMO) is used in *Headar* to achieve higher azimuth-elevation resolution. Superior range and velocity resolutions could be achieved by maximizing the bandwidth swept by single chirp and tuning the chirp interval (i.e., larger chirp interval results in more sensitive velocity measurement but decreased maximum velocity without ambiguity).
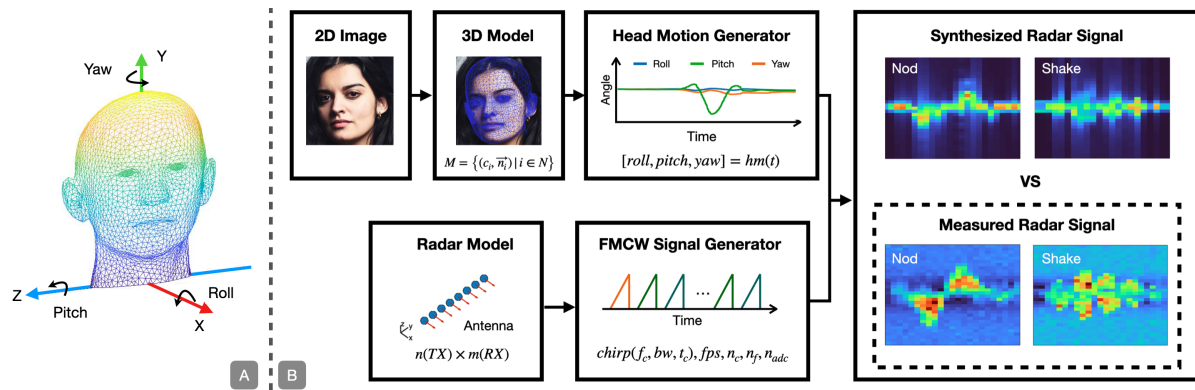
3.2 Simulation

Fig. 2. A: Example of head meshes from P2 (a participant of the study in Section 7). We use the same color convention (i.e., X (red), Y (green) and Z (blue)) throughout the paper to denote coordinates. Rotation around axes (i.e., roll, pitch, yaw) are denoted. B: Overview of our pipeline for synthesizing radar responses to head movements and parameters for each module. The image and its 3D model is obtained from DAD-3DHeads [56] for demonstration purpose in this figure. M: head mesh. $hm(t)$: head posture over time depicted by roll, pitch, yaw. n: number of transmitters. m: number of receivers. f_c: radar operating frequency. bw: chirp sweeping bandwidth. t_c: chirp sweeping duration. n_{adc}: frame per second. n_c: number of chirps per frame. n_f: number of frames. n_{adc}: number of ADC samples per chirp.

3.2.1 Model Radar. To get the radar responses of a user's head during its movement, we modeled a monostatic MIMO radar with three transmitter and four receiver antennas distributed along the Z axis (Figure 2 A). The radar transmits FMCW signals with a center frequency at 77 GHz to the free space, receives and collects signals that are reflected by the head model. The radar and waveform model is implemented with MATLAB Phased Array System Toolbox[3]. For each frame, the radar emits 128 chirps with time division multiplexing (TDM) technique. Note that these configurations are the same as those we use in our hardware system, as we will later discuss in Section 6.1. The received signals are organized into a radar cube, to which an FFT-based fast-time-slow-time processing technique is then applied for range and velocity determination of the head. We then compressed the radar cube by finding the maximum value across the range axis for each Doppler bin, to show the Doppler variation over time.

3.2.2 Model User Head – static. Prior research has modeled the human body as ellipsoid segments to synthesize full-body radar responses [27]. However, in our application, which involves sensing human head at a short range using radar configured to have a relatively high spatial resolution, the variation of radar responses resulting from the differences in head size, shape, facial features etc. among people is significant. Ellipsoid segments would be insufficient to express variances of these head features. Following the conventional approach of generating 3D mesh of human body in a recent work [3], we used a head alignment model DAD-3DHeads[4] to generate a 3D model of a user's head from a single 2D image, and then meshed the 3D model into N triangular patches p_i, as shown in Figure 2 A. Each patch, centered c_i with normal at $\vec{n}_i$, is regarded as a single target responding to incoming radar signals. The signals received by an antenna are calculated as the sum of reflected signals from effective head patches, which will be discussed in Section 3.2.3.

Because these patches vary in size, distance, and visibility to the radar, they yield different reflected signal intensities (aka reflectivity), to quantify which, we used Radar Cross Section (RCS), denoted as σ. RCS depends on the target's orientation (i.e., azimuth, elevation), material, and frequency of incident waves. For simplicity, we modeled a user's head meshes as a surface of a homogeneous material with electromagnetic properties of dry skin, neglecting the effects of internal tissues, bones and different parts of the head such as hair, eyes and lips [10, 18]. According to [18], RCS of human dry-skin can be estimated by $\sigma_{skin} \approx \tau_m \sigma_{PEC}$, where $\tau_m = -4.68dB$ and PEC denotes patches made of perfect electrical conductor reflecting all incident waves. We obtained σ_{skin} by simulating σ_{PEC} of head patches using MATLAB. Due to the short wavelength (3.9 mm) at 77 GHz, the reflectivity of each target is regarded as constant as frequency changes.

3.2.3 Model User Head – dynamic. We animated different head gestures (i.e., *nod* and *shake*) by defining orientations (i.e., roll, pitch and yaw) of the head model as a function of time. We tracked participants performing head gestures in our later survey studies, several examples could be found in Figure 5 (details of data collection can be found in the following sections). We simplified the modeling of a user's head-neck section by using a "ball-stick" model [41, 79] with which the head motion could be depicted with only angular parameters. For each head pose, only head patches visible to the radar can contribute to the reflection of radar signals, thus regarded as effective targets to the radar. We obtained these effective patches from the head model by extracting those with an angle of ± 90 degrees between their normal (i.e., $\vec{n}_i$) and the radar line of sight.

Each patch of the head model has its velocity during head movement. Specifically, when a user performs nodding, the head leans forward and backward and thus shows a negative and positive velocity sequentially from the perspective of radar. However, while shaking, the head rotates around the rest of the human body and has both positive and negative velocity patches at the same time. The velocity of each patch is measured by the radar

---

and induces different radar response patterns, which is the source of signal we hoped to leverage to differentiate the two gestures in *Headar*.

3.2.4 Verification. To prove the effectiveness of the modeling of the radar response, and the user head in static and dynamic settings, we generated a head model of 8,937 patches from an image of P2 (Figure 2 A) and used head gesture motions measured from P2 in the study we later discussed to control the motion of the head model in generating the synthesized radar signals. We plotted a waterfall of maximums of the range bins from the synthesized Range-Doppler Maps and compared it with its corresponding real-world measurements for validation.

Figure 2 B shows the synthesized and measured radar responses for *nod* and *shake* gestures. The synthesized signals have similar features to those of the measured signals for both gestures. Note that we only make a comparison between the dominant signal patterns in the synthesized and measured signals, while ignoring small ambient noise in the background. Specifically, for *nod*, the signals have a sine-wave-like pattern over time, while *shake* yields a sine-wave signal superimposed with its polarity-inverted signal in the axis of Doppler (i.e., velocity). This result confirms the existence of the different motion variances from *nod* and from *shake* as mentioned in Section 3.2.3, and verifies our sensing principle.

4 STUDY OF HOW PEOPLE PERFORM HEAD GESTURES

We investigated the characteristics and variance of users performing head gestures, which informed our development of the sensing pipeline. Additionally, knowing head gestures in *natural* setting yields insights into the design space of *explicit* head gestures, as these *explicit* ones have to be discernable with *natural* head gestures to avoid confusion. For instance, head gestures such as head up, down, left, and right could be used for UI navigation. Specifically, we aim to answer three key questions with this study:

- **Q1**: How many cycles do people *nod* and *shake* their heads when they perform head gestures?
- **Q2**: What is the motion sequence of head gestures?
- **Q3**: What are the duration, magnitude, and speed of head gestures?

4.1 Setup

We conducted a quick investigation in understanding how people perform head gestures, specifically, *nod* and *shake* of people's heads in natural settings (i.e., how people would perform head gestures in daily conversations). We used an OptiTrack system[5] with three groups of markers, with each group constituting a trackable rigid body, attached to a participant's top of the head and two shoulders respectively. Each marker yielded 6 DoF data – position (x,y,z) and orientation (roll, pitch, yaw) in the room frame of reference. Four cameras were set up at the four corners of a $3.2(L) \times 1.9(W) \times 2.2(H)$ m space, and tracked all markers at a resolution of ±0.29 mm with a sampling rate of 120 FPS. Figure 3 A shows this setup.

We recruited 15 participants (with a mean age of 23, 6 females) all of whom were right-handed, usually wearing smartwatches on their left wrists. In the study, participants were asked to perform head gestures following audio commands *nod* and *shake* as naturally as possible – as if they would perform these gestures in daily conversations with others. We instructed participants to look at a reference point while standing naturally before performing each gesture. This served as a quick calibration that allowed us to align the participant's body coordinate with the world coordinate. Each participant performed 10 *nod* and *shake* respectively, in a randomized order to avoid the order effect. We also recorded bio-metric information of participants including body height, arm length, neck length (jaw to clavicle), head length (from head top to chin), and head radius (around forehead). In total, we collected 300 trials of head gestures with 150 for *nod* and *shake* respectively. At the start of each trial, participants stood in a natural position, and the data recorded at this point is considered as the reference point. All subsequent

---

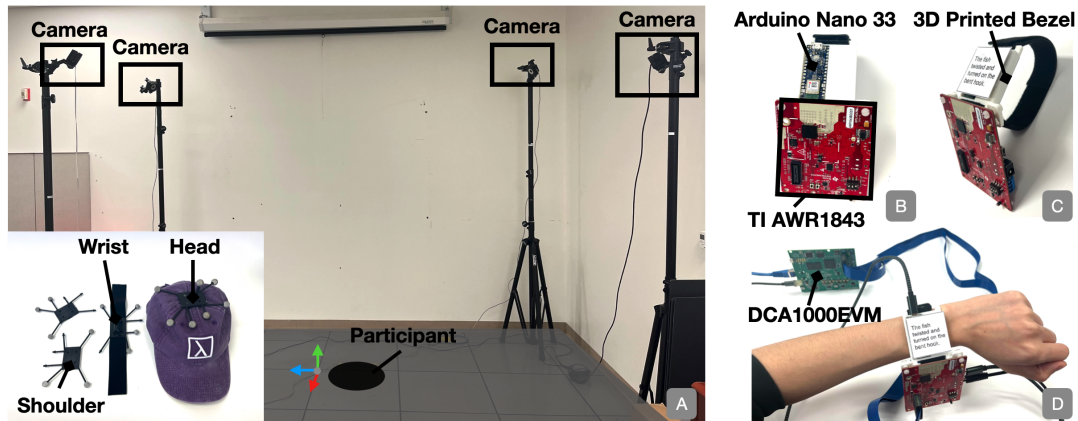[5]OptiTrack Motion Capture: https://optitrack.com/cameras/primex-13/

Fig. 3. *Headar* setup. A: The room configuration and wearble tracking markers for the head gesture study and the smartwatch position/orientation study. The cameras were calibrated to the coordinate represented by the colored axes. B, C: Front view and side view of our system prototype. The radar sensor (i.e., TI AWR1843) is mounted on a 3D-printed bezel. D: Demonstration of a user wearing *Headar*. DCA1000EVM is used to stream data from the radar sensor to a nearby computer for processing.

data points refer to this reference point to determine the participants' relative head position/orientation over time. Figure 4 shows superimposed head orientation of each participant, deposed into roll, pitch, and yaw over time.

4.2 Result

4.2.1 Number of Cycles. Despite the highly accurate 3D spatial tracking, we found algorithm-based cycle detection inevitably yield errors. Alternatively, two experimenters manually examined the plots of head orientation to count the number of cycles. Specifically, for each motion sequence, we omitted peaks and troughs with magnitudes significantly smaller (<10%) than the maximum peak-to-tough magnitude in that sequence. We counted one round trip of a participant's head as a cycle (e.g., head up from neutral and then head down to neutral). A sine-wave alike head trajectory (e.g., Figure 4 P14) would count as *three* cycles. The final count of cycles of each participant was calculated by averaging the two experimenters' results. The result indicated that most participants (n=11) perform head gestures with less than seven cycles for both *nod* and *shake*. On average, participants performed 6.06 cycles (SD=2.32) of head motion in *nod*, and 5.27 (SD=2.06) in *shake*. We investigated variation in one's head gestures by calculating the standard deviations of the number of cycles within each participant's data and found an mean standard deviation of 0.76 (SD=0.41) for *nod*, and 0.77 (SD=0.34) for *shake*. In other words, participants showed some consistency of the number of cycles in their head gestures mostly with fluctuation of less than two cycles. This result undermined the possibility of having heuristic head gesture classifiers using rules based on number of cycles (i.e., peak finding or template matching) for the significant overlap revealed between *nod* and *shake*. The cross-user variation in head motion also exceeded our expectation – some participants performed head gestures with as few as three cycles while there also exist participants who would perform more than ten cycles. We decided to accommodate this variation in designing our signal processing pipeline and training our deep learning models.

4.2.2 Motion Sequence. We also investigated the motion sequence of head gestures. Specifically, we asked: 1) would participants head up or down to initiate a *nod*, and left (counter-clockwise) or right (clockwise) a *shake*? and 2) would participants be consistent across head gestures regarding their their first head motion direction.
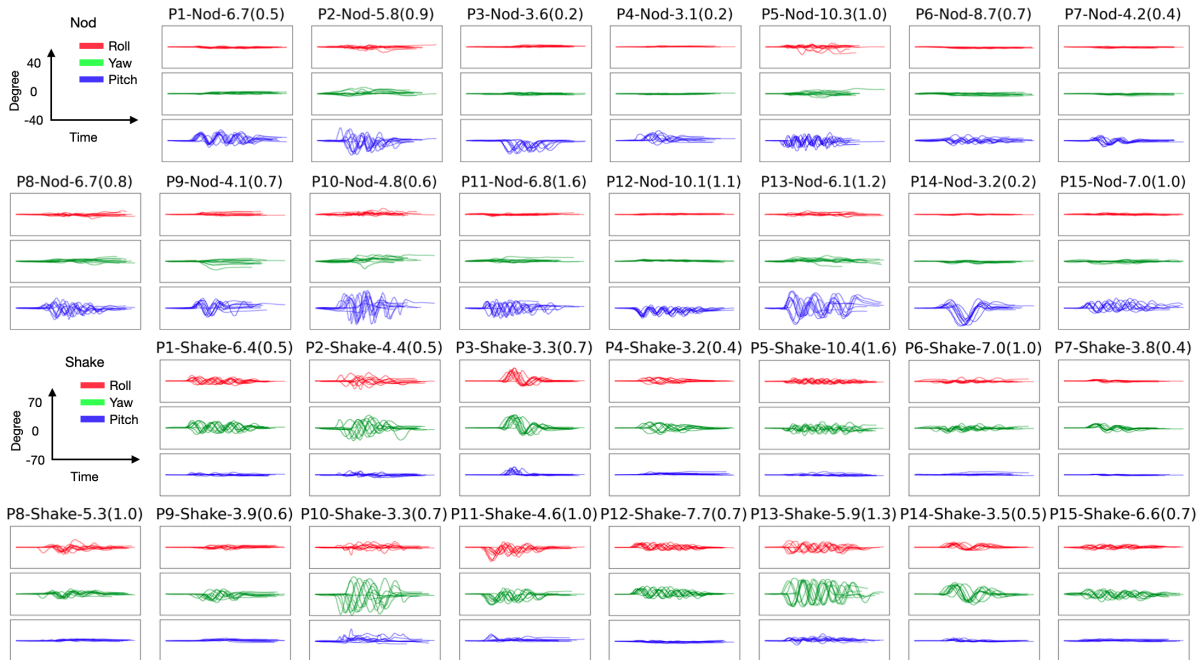
Fig. 4. Superimposed head orientation variance over time of each participant during the course of head gestures, deposed into roll, pitch, and yaw. The label on top of each subplot shows the participant ID, head gesture type, average number of cycles, and standard deviation across trials within a participant's data. The scale is consistent for all plots from each gesture, and is labeled on the legend.

Results of this motion characteristic could yield insights into heuristics that our sensing approach could take advantage of. For example, we found the head-down motion yield negative radial velocity towards a smartwatch radar which is the opposite to the head-up motion. In theory, head left and right should result in conjugated velocities, for half of the user face would always have the opposite velocity as the other half regardless of the head yaw rotation (discussed in Section 3.2.3). However, in practice, the smartwatch radar is rarely positioned aligned with the pivoting line of the head rotation and therefore causing head left and right to have minute differences in signals.

We conducted the same labeling process as used previously and asked the two experimenters to discuss and converge on labels that showed divergence (i.e., only 4 out of 300 labels). We investigated the consistency of participants in initiating head gestures, both individually and across different participants. Results are shown on Figure 5 at the top of each example signal. Out of the 150 head *nods* trajectories we collected, 111 showed head up first while the rest head down first while the rest 39 showing head down first. Among the 150 head *shakes*, 78 showed head left first while the rest head right first, indicating the diversity of motion sequences among people. For *nod*, 8 out of 15 participants never changed the direction of the head direction in initializing their head gestures. It was 4 out of 15 participants for *shake*, in which participants showed less consistency in the direction of initial head motion. The aforementioned results posed challenges in head gesture recognition and eliminate the feasibility of heuristically generated cycle-count- or phase-based algorithms.

4.2.3 Duration. On average, participants took 1.95 seconds (SD=0.63) to complete *nod* head gesture and 2.02 seconds (SD=0.69) for *shake*. Within each participant's data, the mean standard deviation measured 0.30 seconds
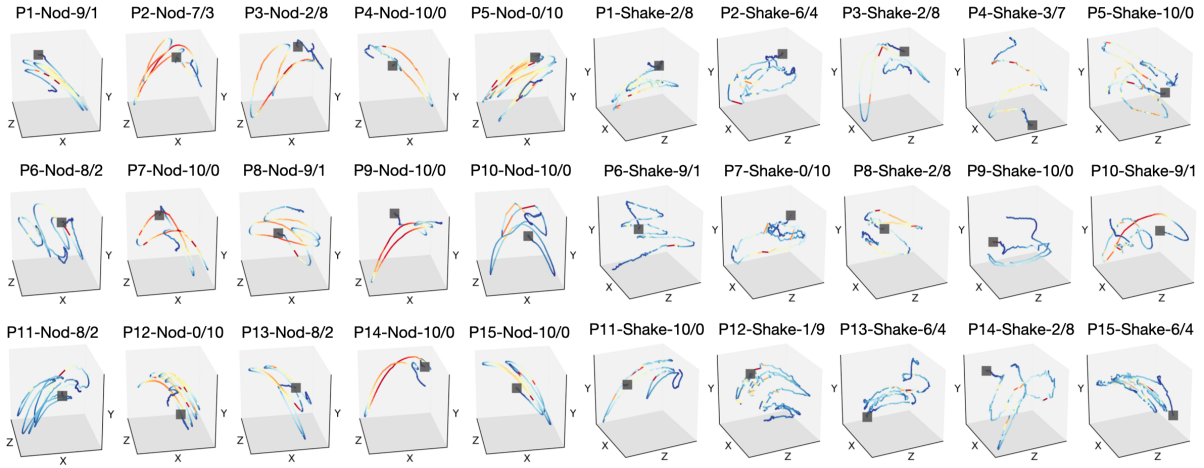
Fig. 5. Example head trajectories of each participant. The label on top of each subplot shows the participant ID, head gesture type, the ratio of the number of head direction for initiating a gesture among the ten trials (i.e., head up/head down for *nod*, head left/head right for *shake*). The original position of participants' head is denoted by a grey square. Velocity of the head motion is color coded (with blue and red denoting low and high velocity). All plots are rendered in the same scale.

(SD=0.11) for *nod* and 0.26 seconds (SD=0.10) for *shake*. Results indicated similarity in duration for both gestures, showing promise in having a signal segmentation of a window of the same length to identify possible radar signals for head gesture recognition. Results from this investigation informed our signal segmentation mechanism which we will detail later in this paper (Section 6.2.2).

4.2.4   Magnitude. We measured the magnitudes of both translational and rotational movements of participants' heads by tracking the motion of the visual marker affixed on the top. The magnitude is defined as the maximum deviation of the head from its reference point. For *nod*, the translational motion magnitudes were measured 7.12 cm (SD=3.29), 1.07 cm (SD=1.12), and 1.67 cm (SD=0.78) for X, Y and Z axes. We also calculated the angular motion magnitude considering the participant's head as a rigid body, from the root of the neck (estimated by the midpoint of the two shoulders) to the top of the head. The angular motion magnitudes measured 2.94 degrees (SD=1.52), 17.20 degrees (SD=6.26), and 3.91 degrees (SD=1.85) for roll, pitch, and yaw respectively. For *shake*, the translational motion magnitudes are 3.36 cm (SD=1.83), 1.37 cm (SD=0.53), and 3.01 cm (SD=1.40) for X, Y and Z axes, and 15.4 degrees (SD=8.19), 7.88 degrees (SD=5.23), and 26.53 degrees (SD=13.42) for roll, pitch and yaw.

4.2.5   Speed. Regardless of the number of cycles, magnitude, or duration, head motion velocity decreases at the turning points where the head changes direction. Figure 5 shows examples of 3D trajectories with velocity coded in color. This distinctive pattern in velocity could be well captured by FMCW velocity sensing on the Range-Doppler map, which we decided to use as our main input for our gesture recognition deep learning pipeline. The trajectories also illustrate that *nod* exhibits a wider range of speed variation compared to *shake*. On average, we measured an averaged linear velocity of 17.52 cm/s (SD=7.48) from a range of 5.54 to 41.72 cm/s across participants from *nod*, 9.68 cm/s (SD=5.51) from a range of 3.18 to 45.62 cm/s from *shake*. For angular velocity, we measured an average of 8.61 degree/s (SD=5.2), 59.14 degree/s (SD=24.80), 7.67 degree/s (SD=3.76) in roll, pitch, yaw across participants from *nod*, and 28.17 degree/s (SD=14.83), 12.26 degree/s (SD=12.89), 84.50 degree/s (SD=36.46) from *shake*. These characteristics informed our hardware configurations which we will discuss in Section 6.1.
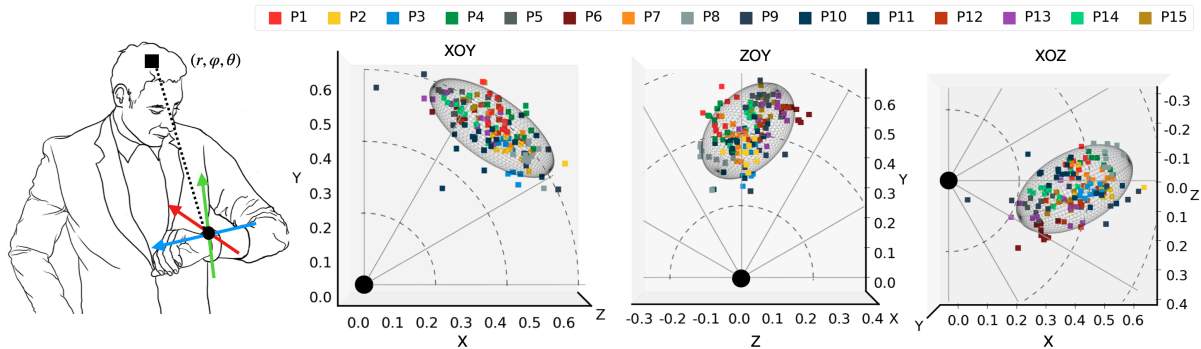
Fig. 6. From left to right: illustration of positioning relationship between heads and smartwatches, projections of aggregated head positions within the watch frame of reference into XOY, ZOY and XOZ planes. The watch is shown in dots. Participants' heads are shown in squares, the color of which denotes the participant ID. A 95% confidence ellipsoid is used to show the region where participants' heads cluster.

5 STUDY OF HOW PEOPLE POSITION AND ORIENTATE SMARTWATCHES

5.1 Setup

We conducted another study investigating how smartwatch users would position the watch with reference to their heads after wrist raise activation, which informed our choices of hardware configuration such as the placement of radar sensors and spatial resolutions, providing valuable insights for future design considerations. Additionally, knowing the head position and orientation allows us to select a region of interest for data analysis, thereby mitigating the impact of background noise.

We used the same group of participants from the previous study. An additional visual marker that yielded 6 DoF of a dummy smartwatch was added to participants' wrists. The dummy smartwatch weighed around 40 g, similar to contemporary smartwatches. In this study, participants were asked to raise the dummy smartwatch up to the position they would use a real smartwatch for interactions (i.e., the wrist-raise-activation position). Each participant was asked to raise the smartwatch ten times. In total, we collected 150 sequences of the smartwatch at the wrist-raise-activation position, each of which lasted around 2 seconds.

5.2 Result

We tracked the head positions in the watch frame of reference. Figure 6 shows the aggregated results from all participants. Among the data we collected, we observed that two data points from P12 were anomalous possibly due to system glitches of OptiTrack, and thus they were precluded in the figure and following statistics.

Participants showed various positions and orientations of their smartwatch after wrist raise activation. Across participants, we measured a mean distance (r) of 63.64 cm (SD=3.08), a mean azimuth angle (φ) of 5.47 degrees (SD=10.86) and a mean elevation angle (θ) of 39.83 degrees (SD=7.22) between the smartwatch and the wearers' heads. This result indicated that azimuth angle showed the largest variance among participants, with watches positioned on both sides around the head axis. Results also showed within-user variance. However, some participants showed more consistent wrist-raise-activation positions than others. Within each participant's data, we measured a mean distance variance of 3.09 cm (SD=1.73), and a mean angular variance of 5.72 degrees (SD=3.36), 5.62 degrees (SD=3.49) for azimuth and elevation respectively.

For these results, we selected a mmWave radar, TI AWR1843, that has a field of view (both azimuth and elevation), sensing range, and resolution sufficient for capturing user head positions assuming a radar instrumentation location on a smartwatch bezel (akin to the sensor instrumentation location of the smartwatch prototype in Google

Soli). Additionally, we also found the magnitude of positional and angular variances needed to be considered for signal processing and to be included in training deep learning models.

6   SYSTEM IMPLEMENTATION

We aim to develop *Headar* for real-time head gesture detection in both stationary and mobile real-world scenarios. We describe the hardware configuration and software design of our system in this section.

6.1   Hardware

Our hardware (Figure 3 B, C, D) is based on the off-the-shelf TI AWR1843BOOST radar sensor, which operates between 77-81GHz [75]. The sensor has three transmitters and four receivers with a 120-degree and 30-degree field of view in the azimuth plane and elevation plane respectively. We configured the radar to transmit 14 frames per second, with 128 chirps in each frame and 256 samples in each chirp. A TI DCA1000EVM adapter [76] is utilized to stream raw ADC data with a Low Voltage Differential Signaling (LVDS) interface from the radar sensor to the computer, where we developed a program to collect UDP packets via Ethernet. The packets are then parsed into data frames for further processing. This pipeline enables a real-time system with a high frame rate, allowing us to observe subtle signals from the head.

To enable a wearable device with the radar sensor, we 3D printed a watch bezel with a size of $44(L) \times 56(W) \times 8(H)$ mm, which is similar to the size of a commercial watch. We mounted the radar sensor to the lower bezel tilting 45 degree upward. This design of placing radar antennas has been seen in Google Pixel Watch[1], considering the fact that the watch display blocks RF signals while plastic materials allow RF penetration. At the base of the bezel we also instrumented an Arduino Nano 33 BLE with an IMU sensor (LSM9DS1) [2], which is common to find on wearables.
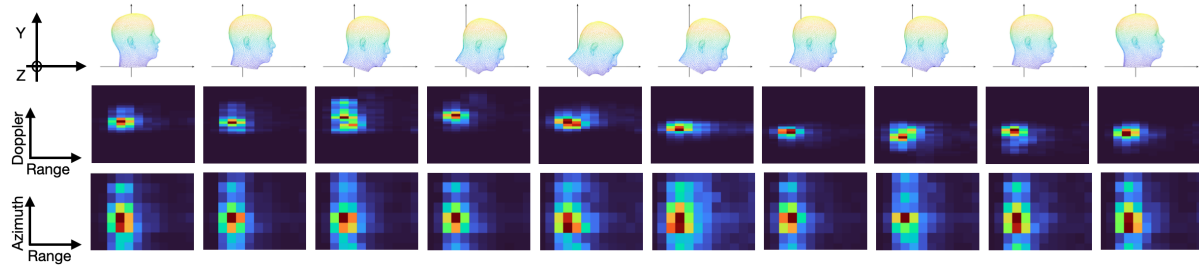

Fig. 7.   Example sequence of *nod*. From top to bottom: side view of head, Range-Doppler, and Range-Azimuth Maps.
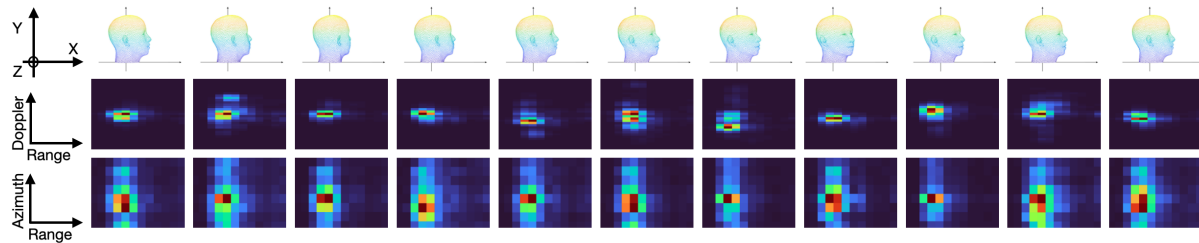

Fig. 8.   Example sequence of *shake*. From top to bottom: side view of head, Range-Doppler, and Range-Azimuth Maps.

6.2 Software

6.2.1 Signal Processing. The raw ADC data is organized into a 256x128x12 radar cube per frame, which is then processed with the same pipeline as discussed in Section 3.2.1 to generate complex Range-Doppler, Range-Azimuth profiles and temporal Doppler sequences. To adapt the pipeline to real-world signals, we made improvements by subtracting the mean value across Doppler to remove static clutter in the environment. To eliminate background noise and reduce unnecessary computational load, we cropped out the data beyond 1 m range and ±1.0 m/s velocity, which results in a data dimension of 12x32x12. Figure 7 and 8 show sequential frames of Range-Doppler and Range-Azimuth Maps for *nod* and *shake*.

6.2.2 Head Gesture Signal Segmentation. To achieve real-time gesture recognition, a segmentation mechanism is developed to obtain effective motion sequences automatically from continuous signals. We refer the start and end of a time window with potential effective motions as "Onset" and "Offset" respectively. We used an IMU sensor to detect a user's wrist status, thus enabling or disabling the radar sensor. Specifically, the radar starts collecting data only when users raise their wrist and stay still. Figure 9 shows onsets and offsets of wrist-raise-activation data collection (i.e., the signal between point#1 and #6), during which *Headar* looks for onsets and offsets of head gestures (i.e., signals between point#2 and #3, and between point#4 and #5). In this example, the mmWave radar was off until point#1 (when the user's wrist is raised), which could be easily detected by the wrist-raise-activation feature on existing smartwatches. A cooling period of 0.7 seconds is implemented to eliminate false positive detection caused by radar signals from wrist-raising motions. Similarly, point#6 is detected by the same feature.

After initiating radar detection, our system keeps monitoring the signals looking for potential head gestures input (e.g., signals between point#2 and #3, and point#4 and #5). Considering the practicability of the wearable system, we eliminated the sliding-window approach for its high computational cost. Instead, we proposed a pipeline featuring dynamic thresholding, based on our observation that the signal strength (SS) during head gestures is ten times greater than when the user is not performing any gestures. Three states are involved in the pipeline: *No Gesture*, *Recording for DL Inference*, and *Waiting for Head Gesture Offset* (Figure 9). Specifically, at each frame, we sum up the Range-Doppler Map to get SS. To trigger gesture "Onset" and "Offset", we threshold SS with $Thred_{on}$ and $Thred_{off}$, which are automatically adjusted by IMU data to lower false positives – both thresholds are set higher when there are larger amplitudes of movements. Once "Onset" is detected (i.e., SS > $Thred_{on}$), we record 30 frames which last roughly 2 seconds before changing the state to *Waiting for Head Gesture Offset*. We chose this time span, based on our study results in Section 4, which revealed that on average people took ~ 2 seconds to complete head gestures. Finally, once offset is detected (i.e., SS < $Thred_{off}$), the system changes to *No Gesture*.

6.2.3 Gesture Existence Detection. Prior work [97] has empowered RF sensing on mobile platform by compensating the phase shift induced by platform motion. We tried a similar approach to enhance the system's ability in mobile settings, but found that movement from a user's torso during walking to be the major source of error rather than the movement of the wrist-instrumented sensor. Due to the sparse spatial information from radar, involuntary head movements during walking, speaking, etc. result in similar signals as those from head gestures with small magnitude. This ambiguity poses challenges to our deep learning inference. To demonstrate this ambiguity, we visualized signals (i.e., *Nod* (P1), *Shake* (P1) and *Walk* (P2)) collected in a later user study in Figure 10. Therefore, we implemented a gesture existence detection layer on the segmented sequence to identify if it contains gestures before feeding positive sequences to deep learning. Specifically, for each frame in the segmented sequence, we examine the variation of signals by normalizing the compressed radar cube, denoted by Var. We threshold Var with $Thred_{var}$ to get rid of sequences with little variations – sequences more likely to contain only involuntary head movements. $Thred_{var}$ is also automatically adjusted by IMU data – the threshold is increased accordingly when the body movement is larger.
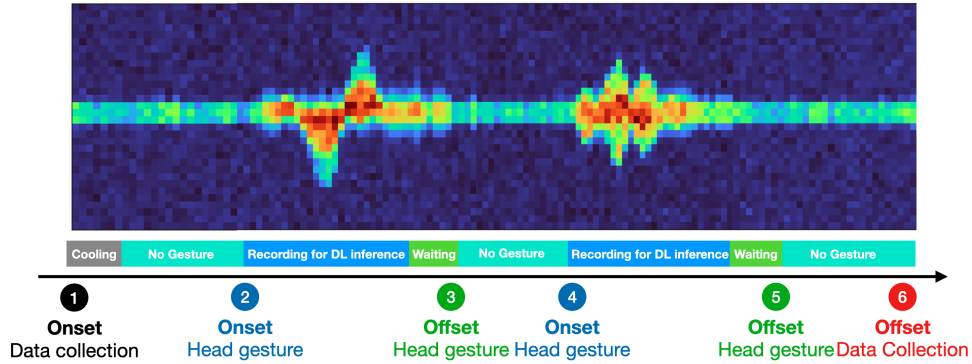
Fig. 9. Illustration of Onset-Offset detection needed in *Headar* in real-world smartwatch uses. From top to bottom: example of a signal sequence including *nod* and *shake*, status of corresponding signal segments, Onsets/Offsets detected with our segmentation algorithm for data collection and head gestures.

6.2.4   Deep Learning Architecture. Prior work [35, 78] has shown encouraging performance of Convolutional Recurrent Neural Network (CRNN) in learning spatial features and Long Short Term Memory (LSTM) in extracting temporal features. Particularly, RadarNet [35] proposed a model for hand gesture recognition, which was proved to be efficient and accurate in practical use by training and testing with a large dataset. We leveraged the state-of-the-art RadarNet architecture in our system. Specifically, a frame model with multiple channels is developed to get spatial features from each frame, which are wrapped around along temporal dimension and then sent to an LSTM layer (32 units) to integrate sequential information. The final feature vectors are then sent to three dense layers to output the classification prediction. The model uses an Adam optimizer with a learning rate of 0.001 and a sparse categorical cross-entropy loss function. The model was trained for 80 epochs with a batch size of 32.

6.2.5   Radar Data Representation and Augmentation. Prior research has investigated various representations of radar signals, with the mainstream approach using complex radar cube [35, 102], Range-Doppler Map (RDM) [65], Range-Azimuth Map (RAM) [80], and Range-Elevation Map (REM) [101]. These representations can be either used individually or combined together as input to construct different feature spaces for deep learning models to learn from. Rather than using a radar cube that contains only implicit information, requiring more training effort to harness, we employed multiple radar profiles (i.e., RDM, RAM, REM) that offer explicit information on the range, velocity, and angles of a user's head. Our user study in the following section investigated the model performance with inputs of different feature spaces. Furthermore, to accommodate different head gesture patterns as we found in Section 4 and improve the robustness of our model, we synthesized head gesture samples by augmenting the dataset collected in the next section. Specifically, to accommodate for the variety of head gesture motion sequences (e.g., head-up-first and head-down-first in *nod*), we inversed motion sequences (e.g., from head-down-first *nod* to head-up-first *nod*) by flipping the radar profiles along the zero-Doppler axis. We also flipped radar profiles along the range axis to reposition the target (i.e., head). Finally, we implemented random scaling to radar profiles to generate more data samples. This data augmentation is applied to the training set in the evaluation we will discuss next.

7   EVALUATION

7.1   Setup

In most use scenarios, users have their body relatively stationary when interacting with smartwatches. We found the difference between performing head gestures while standing and sitting negligible from the measurements

of our radar sensor, and thus in this paper we evaluated the sensing pipeline when users were sitting on a chair to mitigate fatigue. Additionally, we realized that there are situations where users have to interact with smartwatches while in mild motions, which led us to evaluate our sensing approach's ability in mobile settings where a treadmill was configured to 0.6 mile per hour. We mimicked this slow pace of mobility based on our observation that users tend to slow down before interacting with smartwatches for safety purposes – their sights are taken from previous tasks for reading visual content on smartwatch screens. We conducted evaluations in a laboratory setting with participants being stationary or moving with a slow pace. Finally, in our experiments, we did not observe any significant differences in radar measurements between laboratory conditions and common everyday environments. This is because *Headar* only leverages signals from a close-range area around a user's head, effectively excluding most of the surrounding background environments. However, we believe future investigations with a wider range of environments of different characteristics (e.g., EMI noise, temperature, humidity) should be conducted to further our understanding of our proposed sensing approach.

7.2 Procedure

Considering natural head movements and common use cases of smartwatch, users were instructed to react to verbal cues delivered by audio clips. In this study, we collected nine motions including *Nod*, *Shake*, *Speak*, *Touch*, *Still*, *Head Up*, *Head Down*, *Head Left* and *Head Right*. Specifically, we instructed participants to *Nod* and *Shake* their heads facing the worn smartwatch prototype (Figure 3 D) at the wrist-raise-activation position, using the same manner as they would do in daily communications. *Touch* motion involved participants touching four buttons printed on the paper smartwatch face (Figure 3 C). In the *Speak* motion, participants were asked to read a sentence from the Harvard Sentences[6] commonly used to test speech quality in telecommunication for their balanced phonemes. One sentence was printed on the paper smartwatch face, and was randomly selected for each participant. In the *Still* condition, participants were requested to look at their smartwatches while holding it to the wrist-raise-activation position without performing any motion explicitly. In the rest four orientational head motions, participants were asked to turn their head up/down/left/right and stay in the position until they heard the instructions from the experimenter to turn their head back.

We recruited 15 participants (with a mean age of 23, 6 females) for the evaluation. For each participant, there were 8 data collection sessions in the sitting scenario and 2 sessions in the walking scenario. Within one session, the number of the aforementioned nine motions is 8, 8, 1, 1, 2, 1, 1, 1, 1 accordingly, counting 24 motions in total. This split on the number of trials yielded balanced datasets that consist equal amount of *Nod* and *Shake* head gestures, and the other label including the rest 7 motions (i.e., *Touch*, *Still*, *Speak*, *Head Up/Down/Left/Right*). The verbal cue was initiated by key pressing by the experimenter and the commands were delivered in random order. Once the start key was pressed, we started to record data from the mmWave radar. This data included the cropped radar cube as discussed in Section 6.2.1, from which RDM, RAM and REM can be derived. We configured both the mmWave radar and the IMU sensor to stream data at 14 frames per second. The experimenter stopped the data streaming by pressing the stop key after visually confirming that participants had finished the requested motions. At the end of each session, data was saved to files for further processing. In total, we collected 1200 trials of *Nod* and *Shake* respectively, 300 trials of *Still*, 150 trials of each of the rest six motions. We augmented the data samples with the approach discussed in Section 6.2.5 and got 10800 samples in total. We then trained our deep learning model with a structure discussed in Section 6.2.4 and evaluated the model.

7.3 Result

We ran a Leave-One-Session-Out (LOSO) validation within the ten sessions for each participant. Specifically, we ran a one-fold cross test among the ten sessions by using one session as the test set and the rest nine sessions of

---

[6]Harvard Sentences: https://en.wikipedia.org/wiki/Harvard_sentences

data as the train set. Overall we got a 85.5% (SD=11.1) accuracy among all users and the aggregated confusion matrix is shown in Figure 11 (left). It is expected that for a particular participant, a repetitive pattern can be observed in this participant's gestures belonging to the same category. To evaluate our model's ability to migrate to unseen users, we ran a Leave-One-User-Out (LOUO) validation across users. All of the ten sessions of data from the left out participant were used for testing and data from the rest 14 participants were used for training. We ran the LOUO for each participant and the accuracy is 82.5% (SD=12) on average. Figure 11 (middle) shows the confusion matrix.
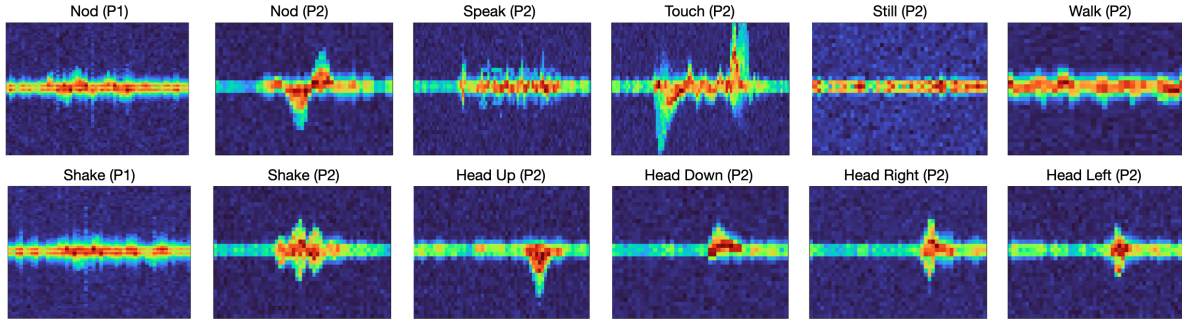

Fig. 10. Examples of radar Doppler sequences for head gestures and other motions from P1 and P2 (participants of the study).
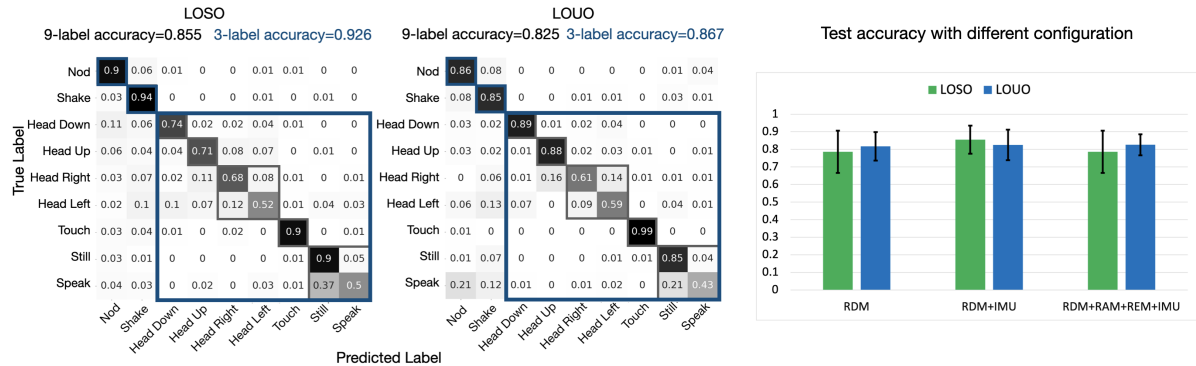

Fig. 11. Left, middle, right: confusion matrix of the LOSO (within-user) validation, confusion matrix of the LOUO (cross-user) validation, test accuracy of model trained with data in different feature spaces (i.e., RDM, RDM+IMU, RDM+RAM+REM+IMU) in LOSO and LOUO validation.

To tease out the performance and sensing capabilities of *Headar*, we conducted a series of ad-hoc data analysis and an additional series of studies as follows.

7.3.1 Nod-Shake-Other 3-Label Bundling.
Considering real-world uses of *Headar* in which confirmation dialogs would discard all gestures other than *nod* and *shake*, we bundled the seven labels on the confusion matrices (Figure 11 dark blue rectangles) so that confusions within the bundled label set were not considered as error – in real-world uses they would simply be ignored the same way for not being head gestures. This bundling increased the accuracy by 7.1% and 4.2%, resulting in an accuracy of 92.6% and of 86.7% for LOSO and LOUO respectively. This improvement was due to the similarity and thus confusion among some labels within the other label set. For

example, *Head Left* and *Head Right* motions yield similar range Doppler responses. Similarity in signals can also be found between *Still* and *Speak*.

7.3.2 Composition of Feature Spaces. As mentioned in Section 6.2.5, various radar profiles can be utilized to create feature spaces. In this section we investigated how leveraging different information could affect the performance of *Headar*. Besides different radar profiles, we also examined the incorporation of IMU data as input to our deep learning pipeline. This idea of using IMU signals was inspired by our observation that wrist (i.e., radar sensor) motions are one of the main sources of noise that manifest on radar measurements and can induce detection errors. By feeding IMU signals as an input entirely consisting of noise (i.e., no signals from a user's head), we could train a deep learning model to better separate noise from signals. Of note that we also tried conventional approaches (e.g., adaptive noise cancellation [54, 90]) but none achieved performance as promising as deep learning. Specifically, we compared performance of *Headar* with models trained and tested by three types of inputs – an input considering only RDM, and inputs the involve RDM+IMU, RDM+RAM+REM+IMU. Figure 11 (right) shows the accuracies of LOSO nad LOUO validation with processes same as ones described previously in Section 7.3. Results indicated that for both LOSO and LOUO validation, including IMU data slightly improved the model performance. Adding additional channels of RAM and REM showed improvement in LOUO but deterioration in LOSO. We think this is because RAM and REM introduce the variation of head and wrist position among different users to the model, thus improving the performance in LOUO. However, these information may confuse the model in LOSO, in which the variation of position is relatively small for single user.

7.3.3 Possibility of Having Robust User-Independent DL Models. Results indicated promise in having a user-independent head gesture recognition pipeline, however, its performance largely depended on the generalizability of deep learning models. We investigated this by investigating the two participants that showed the worst accuracies in the LOUO validations without data augmentation – P6 and P13 with accuracies of 45.3% and 35.0% respectively. These two participants showed the largest improvement after applying data augmentation to increase the variety of training data (with accuracies of 72.4% and 53.63%). They also showed the largest improvements when using user-dependent DL models (with accuracies of 80.2% and 63.5%). Further analysis revealed insights that explained the sources of error. P6 showed particularity in terms of the head motion sequence. P6 tended to raise their head by quite a significant angle to initiate the nodding. This was different from the rest of the participants and led to 64% of P6's head nods being misclassified. After applying data augmentation, in which we paid efforts on introducing more variety in head motion sequence, P6's head nods were 95% classified. For P13, we realized that the angle of the smartwatch was oriented differently than the rest of the participants, with most of their head falling outside the elevational field of view of the radar sensor. This resulted in a significantly less RF reflection and thus a lower Signal-to-Noise Ratio (SNR) than the rest of the participants. We suspect that both cases could be improved by recruiting a larger group of participants yielding higher diversity in training that could lead to more robust user-independent DL models.

7.3.4 Bio-metric Factors on Sensing Performance. We realized that hair has an impact on radar signals due to the short wavelength and the short sensing distance. Long hair can yield noise in radar signals due to its random movement during head motion. Overall, the sensing performance for participants with short hair is better (86.8% in LOSO, 86.5% in LOUO) than that for participants with long hair (78.8% in LOSO, 76.3% in LOUO). This can be improved by including more participants with more hairstyles in training in future work.

We also conducted a one-way ANOVA analysis to investigate the significance of the effects from bio-metric factors on our sensing performance. Results (Table 1) showed that body height, neck length and arm length have no impact observed on the sensing performance, while head radius, head length and gender is correlated with the sensing performance, with P-value of lower than 0.05. This is consistent with our expectation, since different

head features can yield different radar responses as discussed in Section 3. Gender shows a significant impact mainly due to the difference in hairstyles as aforementioned.

Table 1. Table for P value and F value (*: P < 0.05).

	Body Height	Head Radius	Head Length	Neck Length	Arm Length	Gender
Within-User	P = 0.38	P = 0.03*	P = 0.01*	P = 0.58	P = 0.08	P = 0.02*
	F = 3.73	F = 5.85	F = 13.5	F = 0.83	F = 4.57	F = 6.47
Cross-User	p = 0.32	P = 0.03*	P = 0.06	P = 0.07	P = 0.31	P = 0.08
	F = 5.61	F = 6.14	F = 5.16	F = 3.04	F = 1.74	F = 3.74

7.3.5 Head Tracking. With *Headar*, it is also possible to track a user's head using mmWave radar. We aggregated RAM and REM from all participants in the *Still* condition. Figure 12 shows the results which indicated that the mmWave radar we used often had the user head within its field of view with it positioned around 10-30 degrees of the radar's principle sensing axis. User's head could be well captured in both RAM and REM. With this data, we could 1) sense the presence of user's head in the field of view of the mmWave radar, and 2) leverage its location to calibrate for variance in smartwatch wrist-raise-activation positions.
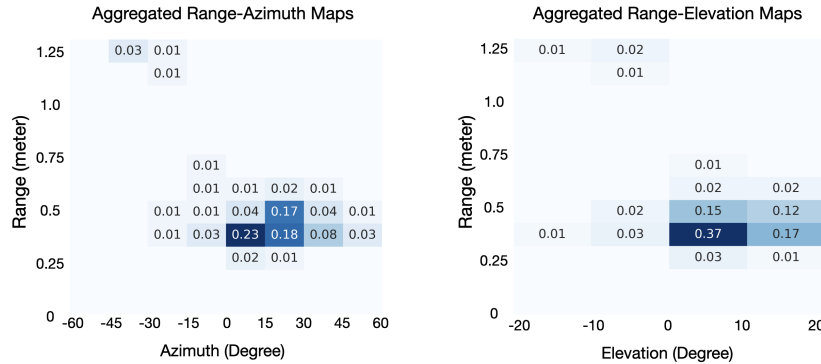

Fig. 12. Left: Aggregated Range-Azimuth Maps. Right: Range-Elevation Maps. Both are from all participants in the *Still* condition. The number represents the probability of users' heads appearing at a specific spatial location.

7.4 Live Test with Unseen Users

We have also conducted a live test with 8 participants (with a mean age of 21, 3 females). These participants were not involved in previous data collection and thus are "unseen users" to the model trained in our previous user study. Our detection pipeline remained unchanged throughout the live test, with no adjustments made to parameters across participants. Participants kept their facial features as they usually do (i.e., two wore a mask, three wore glasses). We used a similar setup with the previous study. Specifically, each participant spent 10 minutes playing around with the device and learning about the uses of our sensing pipeline, and then was requested to complete 18 rounds of live tests in a stationary and mobile setting respectively. The test setup was the same as that of the previous user study (Section 7.1). To mimic the real-world uses of smartwatches, one complete round included raising wrist, performing one gesture per the instruction from an audio command, and resting the wrist. The audio commands included 6 *Nod*, 6 *Shake*, 2 *Touch*, 2 *Speak*, and 2 *Still* in a randomized

order, which were initiated by the experimenter by key presses. We only recorded the live detection result for analysis.

Considering practical uses of confirmation dialogs which often consist of three labels – *Nod*, *Shake*, and *Others* (i.e., neutral), we calculated the accuracy from this evaluation by bundling the labels (i.e., *Speak*, *Touch*, and *Still* were counted as *Others*), which was the same bundling process as the one shown in Section 7.3). Overall, results indicated an accuracy of 72.6% (SD=6) among all participants. We reveal our interpretation of this result as well as findings from the study as follows.

7.4.1 Performance Comparison in Stationary and Mobile Settings.
We have adopted strategies such as dynamic thresholding and IMU-included feature representation to enhance our system's performance in mobile settings (Section 6). Our techniques have been proven effective as evidenced by the comparable performance of *Headar* in sitting and mobile scenarios, with accuracies of 73.6% (SD=6) and 71.5% (SD=5) respectively. The accuracy of gesture segmentation (e.g., point#2 and #3 in Figure 9) is 97.6% in *Headar*, with 7 out of 32 *Still* being misinterpreted as potential gesture input in the walking scenario. All gestures were well segmented in stationary settings. Our system achieved almost equal accuracies for two of the eight users in both settings, and higher accuracy for one user in mobile setting than in stationary setting, indicating that our system has a certain ability to cope with human body motion (i.e., noise). Overall, this showed promise in head gesture sensing with wireless signals in mobile settings.

7.4.2 User Adaptation to the Sensor.
The performance of *Headar* highly depends on the orientation and distance between the sensor and the head, despite our attempts to mitigate this issue and enhance *Headar*'s adaptability to users by incorporating RAM and REM in the model training process, which did not result in much improvement. Upon instrumentation, participants might struggle to achieve good detection results at first until they found the optimal position of the sensor and learned to perform distinctive head gestures after several trials. The detection accuracy improved quickly after participants familiar themselves with the sensing performance. This user adaptation is one of the limitations of *Headar*. We found this adaptation a common practice for emerging interaction techniques. For example, users adapt their face positions and orientations for faster iPhone unlocking using FaceID; users adjust their accent when using speech recognition systems to improve recognition accuracy. Similar to these examples, users could adapt to *Headar* by yielding head gestures with optimal head position/orientation, magnitude of movements and numbers of cycles in their daily uses.

Though the study of usability considering user acceptance and experience to *Headar* was beyond the scope of this work, future research is necessary to enhance the system's ability to generalize to diverse users and environments, including a careful design of wearable radar sensors, refinement of signal processing algorithms and deep learning models, and augmentation of the dataset with a larger and more diverse sample of users. We also envision that a radar signal visualization could be implemented on smartwatches to assist users' adaptation of their head gestures for using *Headar*. It was our observation that visualizations of radar signals to users helped them get a sense of effective signals quickly and expedite the learning process.

7.4.3 Trade-off between Sensitivity and Robustness.
We found that the radar signals are very sensitive to minor movement. Head gestures with small magnitude can be detected with *Headar* but rates of false positive detection also increase, since involuntary head and body movement is inevitable when smartwatches are positioned at the wrist-raise-activation position. For example, participants may move their head and body while speaking and walking. Therefore, we tuned the system to be less sensitive by increasing the thresholds for the detection of onsets and signal existence to improve its robustness. As a trade-off, *Headar* in real-world uses may require users to perform head gestures with a more generous magnitude if they usually perform head gestures very slightly or when there exists significant ambient movement. This adjustment is not significant and is a common experience in our daily lives to cope with a noisy environment. For example, users increase their speaking volume when

using a speech recognition system in a noisy environment. Despite the trade-off, most users in the live test got used to the sensor after only a few trials.

8 DISCUSSION

8.1 Comparison with Prior Work

With the task of head gesture classifications similar to ours, prior studies have primarily employed IMU sensors embedded in head-mounted devices such as earphones [12], VR headsets [32, 79] and eyeglasses [57, 93], or used camera [38, 56, 59, 68] and motion tracking system [77] to detect head movement. These previous work have achieved high accuracy (>93%) in detecting multiple head gestures attributed to the utilization of IMU sensors which can capture explicit head movement information, and the state-of-the-art face recognition techniques which have benefited from extensive training with a large-scale image dataset. Despite their great performance, none of these techniques can be implemented on smartwatches due to the IMU sensor's placement and privacy concerns about using camera.

Closer to our work are those that leveraged mmWave radar to detect head movements [14, 16]. The work by Raja et al. [65] explored the configuration of radar antennas and Doppler effect induced by head movements, and obtained a maximum accuracy of 92% in in-vehicle environments. These works offered preliminary results that demonstrate the feasibility of mmWave-based head gesture detection without evaluating a wider range of real-world scenarios and head movements. In comparison, *Headar* achieved an average accuracy of 84.0% across 9 head gestures, and 72.6% in a live test (i.e., 3 classes with 5 gestures) with a wearable sensor form factor geared towards smartwatch uses, for which our end-to-end pipeline provides a more comprehensive solution. Indeed, as more research emerges showing the exceptional capability of mmWave radar in detecting subtle movements such as hand gestures [50, 78], real-time responsiveness and user independence are gathering increasing attention from researchers for the development of a practical system [35, 52]. The investigation of *Headar* embraced these factors and achieved comparable performance despite future work is needed for further improvements.

8.2 Data Synthesis

Our simulation could be further developed for data synthesis to train deep learning models in the future. Improvements are needed. Instead of a single layer dry-skin model, a more realistic representation of the human head could be achieved by adopting a multi-layer model that incorporates the electromagnetic properties of tissues, bones, hair, and eyes. Prior research has investigated these properties [62, 94], and incorporating their findings could lead to greater precision in radar data synthesis. Additionally, it is also needed to model the radar signal propagation more accurately to better reflect real-world environments [24, 55].

This work only employed the radar data synthesis pipeline to validate the understanding of the sensing principle. However, this pipeline has the potential to synthesize data to improve adoption of radar technology in practice, which has been limited by the lack of large datasets for training. As opposed to prior work that use only range and velocity information of target to synthesize radar data [3], our pipeline includes the modeling of both the radar and the target with higher fidelity, which could potentially lead to flexible and efficient generation of radar datasets for various applications.

8.3 Facial Features on Sensing Performance

In our study, we included participants with different hairstyles including mid/long unstrapped hair, and face wears such as glasses and masks. We are cautious regarding robustness against a wider spectrum of facial features such as skin tones and facial hairs which have been posing challenges in conventional vision-based sensing but should not affect our RF based sensing performance significantly. We suspect that significant factors would include features such as bio compositions, skin moisture, ear drops, and scarfs. We acknowledge that user sets

of a wider diversity should be investigated in the future to further our understanding of the proposed sensing approach.

8.4 Custom Hardware

We used commercially available sensors for ease of development and replication. However, we speculate a performance improvement with custom hardware and fine-tuned signal parameters, in particular, radar antenna laytout and front-end designs optimized for short-range and low-velocity application of gestural interactions. Similar improvements have been shown in Project Soli [50, 78]. Specific directions for improvement include 1) improving the range resolution by having larger frequency-sweep bandwidth, and 2) increasing the field of view to accommodate smartwatch uses at extreme positions/orientations. Finally, the location of the antenna could be optimized using insights from the user study (Section 5). With fabrication techniques to implement soft electrodes (e.g., [83]), antennas could be implemented on watchstraps for better position/orientation that more likely to have a user's head inside the field of view. It might also be possible to implement multiple antennas and use software to select the optimal one for sensing, accommodating a wider variety of smartwatch uses.

8.5 Stress Test under Vibratory Environments and with User Body in Motion

This work did not include a stress test with noise from vibratory environments such as bikes, buses, airplanes, and body in extreme motion conditions such as jogging and running. These sources of noise should be considered in our future work to further our understanding of the performance envelope of the proposed sensing approach. We also speculate that the IMU sensor, which is ubiquitous on smartwatches could show more merits in mitigating vibratory and motion noise than it showed in the current study (Section 7.3) in which only mild motions were induced while participants were walking slowly on a treadmill. Further investigation is needed.

8.6 Richer Set of Gestures

Headar allows *nod* and *shake* as natural head gestures to power confirmation dialogs in a wide array of use scenarios of smartwatches. In addition to the two head gestures *Headar* detects, we also envision future systems to enable a wider array of head gestures for richer set of applications. For example, despite the similarity between head motions (head up, down, left, and right), we showed promise in differentiating between them, and from *nod* and *shake*. Such head gestures could be used for navigating through a list of items on wearable devices to mitigate the "fat finger" problem. Additionally, with the smartwatch being able to sense if its wearer is speaking, we could use future systems to address the "cocktail" party problem when multiple concurrently sources of speech present in the environment after the smartwatch is raised to the wrist-raise-activation position.

9 CONCLUSION

We presented *Headar*, a wearable mmWave radar sensing approach for smartwatches to recognize user head gestures to establish natural interactions for confirmation dialogs. Smartwatches are smarter than ever for their rich activity sensing capabilities and versatile deep learning models, and rely on confirmation dialogs to loop in users in the semi-automation process. We developed and evaluated our system with a series of modeling, simulations, and user studies, the results of which indicated the feasibility of our approach. We envision *Headar* to have a broad impact on smartwatch implementation and interaction designs for the promise and practicality shown in this research, being widely used in designing and prototyping wearable head gesture sensing, while creating a foothold for future work.

ACKNOWLEDGMENTS

REFERENCES

[1] 2006. IEEE Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 kHz to 300 GHz. *IEEE Std C95.1-2005 (Revision of IEEE Std C95.1-1991)* (2006), 1–238. https://doi.org/10.1109/IEEESTD.2006.99501

[2] 2023. Arduino Nano 33 BLE. https://docs.arduino.cc/resources/datasheets/ABX00030-datasheet.pdf Retrieved July 17, 2023.

[3] Karan Ahuja, Yue Jiang, Mayank Goel, and Chris Harrison. 2021. Vid2Doppler: Synthesizing Doppler Radar Data from Videos for Training Privacy-Preserving Activity Recognition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 292, 10 pages. https://doi.org/10.1145/3411764.3445138

[4] Karan Ahuja, Sven Mayer, Mayank Goel, and Chris Harrison. 2021. Pose-on-the-Go: Approximating User Pose with Smartphone Sensor Fusion and Inverse Kinematics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 9, 12 pages. https://doi.org/10.1145/3411764.3445582

[5] Karan Ahuja, Vivian Shen, Cathy Mengying Fang, Nathan Riopelle, Andy Kong, and Chris Harrison. 2022. ControllerPose: Inside-Out Body Capture with VR Controller Cameras. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 108, 13 pages. https://doi.org/10.1145/3491102.3502105

[6] Deepak Akkil, Jari Kangas, Jussi Rantala, Poika Isokoski, Oleg Spakov, and Roope Raisamo. 2015. Glance Awareness and Gaze Interaction in Smartwatches. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI EA '15)*. Association for Computing Machinery, New York, NY, USA, 1271–1276. https://doi.org/10.1145/2702613.2732816

[7] Android Developers. 2022. Dialogs. https://developer.android.com/training/wearables/compose/dialogs Retrieved November 07, 2022.

[8] Apple. 2022. Alerts. https://developer.apple.com/design/human-interface-guidelines/components/presentation/alerts/ Retrieved November 07, 2022.

[9] Homa Arab, Lydia Chioukh, Mansoor Dashti Ardakani, Steven Dufour, and Serioja Ovidiu Tatu. 2020. Early-stage detection of melanoma skin cancer using contactless millimeter-wave sensors. *IEEE Sensors Journal* 20, 13 (2020), 7310–7317.

[10] Domenic Belgiovane, Chi-Chih Chen, Ming Chen, Stanley Y-P Chien, and Rini Sherony. 2014. 77 GHz radar scattering properties of pedestrians. In *2014 IEEE Radar Conference*. IEEE, 0735–0738.

[11] Sejal Bhalla, Mayank Goel, and Rushil Khurana. 2022. IMU2Doppler: Cross-Modal Domain Adaptation for Doppler-Based Activity Recognition Using IMU Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 145 (dec 2022), 20 pages. https://doi.org/10.1145/3494994

[12] Hongliang Bi and Jiajia Liu. 2022. CSEar: Metalearning for Head Gesture Recognition Using Earphones in Internet of Healthcare Things. *IEEE Internet of Things Journal* 9, 22 (2022), 23176–23187.

[13] Mehdi Boukhechba, Lihua Cai, Congyu Wu, and Laura E Barnes. 2019. ActiPPG: using deep neural networks for activity recognition from wrist-worn photoplethysmography (PPG) sensors. *Smart Health* 14 (2019), 100082.

[14] Drew Bresnahan and Yang Li. 2021. Driver head motion monitoring using a mm-wave FMCW radar. In *2021 IEEE Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)*. IEEE, 1–4.

[15] Fabian Bross. 2020. Why do we shake our heads?: On the origin of the headshake. *Gesture* 19, 2-3 (2020), 269–298.

[16] Rachel Chae, Anna Wang, and Changzhi Li. 2019. FMCW radar driver head motion monitoring based on Doppler spectrogram and range-Doppler evolution. In *2019 IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet)*. IEEE, 1–4.

[17] Dongyao Chen, Mingke Wang, Chenxi He, Qing Luo, Yasha Iravantchi, Alanson Sample, Kang G. Shin, and Xinbing Wang. 2021. *MagX: Wearable, Untethered Hands Tracking with Passive Magnets*. Association for Computing Machinery, New York, NY, USA, 269–282. https://doi.org/10.1145/3447993.3483260

[18] Ming Chen, Chi-Chih Chen, Stanley Y-P Chien, and Rini Sherony. 2014. Artificial skin for 76–77 GHz radar mannequins. *IEEE Transactions on Antennas and Propagation* 62, 11 (2014), 5671–5679.

[19] Tuochao Chen, Yaxuan Li, Songyun Tao, Hyunchul Lim, Mose Sakashita, Ruidong Zhang, Francois Guimbretiere, and Cheng Zhang. 2021. NeckFace: Continuously Tracking Full Facial Expressions on Neck-Mounted Wearables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 58 (jun 2021), 31 pages. https://doi.org/10.1145/3463511

[20] Tuochao Chen, Benjamin Steeper, Kinan Alsheikh, Songyun Tao, François Guimbretière, and Cheng Zhang. 2020. C-Face: Continuously Reconstructing Facial Expressions by Deep Learning Contours of the Face with Ear-Mounted Miniature Cameras. 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, 112–125. https://doi.org/10.1145/3379337.3415879

[21] Dale Wilson. 2022. Basics of Millimeter Wave Technology. https://www.allaboutcircuits.com/technical-articles/basics-of-millimeter-wave-mmwave-technology/ Retrieved November 07, 2022.

[22] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software amp; Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 457–466. https://doi.org/10.1145/2807442.2807499

[23] Shinya Fujie, Yasuhi Ejiri, Kei Nakajima, Yosuke Matsusaka, and Tetsunori Kobayashi. 2004. A conversation robot using head gesture recognition as para-linguistic information. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)*. IEEE, 159–164.

[24] F Fuschini, S Häfner, M Zoli, R Müller, EM Vitucci, D Dupleich, M Barbiroli, J Luo, E Schulz, V Degli-Esposti, et al. 2017. Analysis of in-room mm-Wave propagation: Directional channel measurements and ray tracing simulations. *Journal of Infrared, Millimeter, and Terahertz Waves* 38 (2017), 727–744.

[25] O.P. Gandhi and A. Riazi. 1986. Absorption of Millimeter Waves by Human Beings and its Biological Implications. *IEEE Transactions on Microwave Theory and Techniques* 34, 2 (1986), 228–235. https://doi.org/10.1109/TMTT.1986.1133316

[26] Xiangyu Gao, Guanbin Xing, Sumit Roy, and Hui Liu. 2019. Experiments with mmwave automotive radar test-bed. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 1–6.

[27] Andrew Gigie, Smriti Rani, Arijit Chowdhury, Tapas Chakravarty, and Arpan Pal. 2019. An Agile Approach for Human Gesture Detection Using Synthetic Radar Data. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) *(UbiComp/ISWC '19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 558–564. https://doi.org/10.1145/3341162.3349332

[28] Jun Gong, Zheer Xu, Qifan Guo, Teddy Seyed, Xiang 'Anthony' Chen, Xiaojun Bi, and Xing-Dong Yang. 2018. WrisText: One-Handed Text Entry on Smartwatch Using Wrist Gestures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173755

[29] Jun Gong, Xing-Dong Yang, and Pourang Irani. 2016. WristWhirl: One-Handed Continuous Smartwatch Input Using Wrist Gestures. 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, 861–872. https://doi.org/10.1145/2984511.2984563

[30] Anhong Guo and Tim Paek. 2016. Exploring Tilt for No-Touch, Wrist-Only Interactions on Smartwatches. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 17–28. https://doi.org/10.1145/2935334.2935345

[31] Unsoo Ha, Sohrab Madani, and Fadel Adib. 2021. WiStress: Contactless stress monitoring using wireless signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–37.

[32] Tomasz Hachaj and Marcin Piekarczyk. 2019. Evaluation of pattern recognition methods for head gesture-based interface of a virtual reality helmet equipped with a single IMU sensor. *Sensors* 19, 24 (2019), 5408.

[33] John Paulin Hansen, Florian Biermann, Janus Askø Madsen, Morten Jonassen, Haakon Lund, Javier San Agustin, and Sebastian Sztuk. 2015. A Gaze Interactive Textual Smartwatch Interface. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (Osaka, Japan) *(UbiComp/ISWC'15 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 839–847. https://doi.org/10.1145/2800835.2804332

[34] Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: Appropriating the Body as an Input Surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 453–462. https://doi.org/10.1145/1753326.1753394

[35] Eiji Hayashi, Jaime Lien, Nicholas Gillian, Leonardo Giusti, Dave Weber, Jin Yamanaka, Lauren Bedal, and Ivan Poupyrev. 2021. RadarNet: Efficient Gesture Recognition Technique Utilizing a Miniature Radar Sensor. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 5, 14 pages. https://doi.org/10.1145/3411764.3445367

[36] Seongkook Heo, Michelle Annett, Benjamin J Lafreniere, Tovi Grossman, and George W Fitzmaurice. 2017. No Need to Stop What You're Doing: Exploring No-Handed Smartwatch Interaction.. In *Graphics Interface*. 107–114.

[37] Jonggi Hong, Seongkook Heo, Poika Isokoski, and Geehyuk Lee. 2015. SplitBoard: A simple split soft keyboard for wristwatch-sized touch screens. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1233–1236.

[38] Sebastian Hueber, Christian Cherek, Philipp Wacker, Jan Borchers, and Simon Voelker. 2020. Headbang: Using Head Gestures to Trigger Discrete Actions on Mobile Devices. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 17, 10 pages. https://doi.org/10.1145/3379503.3403538

[39] Jessi E. Johnson, Oliver Shay, Chris Kim, and Catherine Liao. 2019. Wearable Millimeter-Wave Device for Contactless Measurement of Arterial Pulses. *IEEE Transactions on Biomedical Circuits and Systems* 13, 6 (2019), 1525–1534. https://doi.org/10.1109/TBCAS.2019.

2948581

[40] Ashish Kapoor and Rosalind W Picard. 2001. A real-time head nod and shake detector. In *Proceedings of the 2001 workshop on Perceptive user interfaces*. 1–5.

[41] Duane V Knudson and DV Knudson. 2007. *Fundamentals of biomechanics*. Vol. 183. Springer.

[42] Gierad Laput and Chris Harrison. 2019. Sensing fine-grained hand activity with smartwatches. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[43] Gierad Laput, Robert Xiao, Xiang 'Anthony' Chen, Scott E. Hudson, and Chris Harrison. 2014. Skin Buttons: Cheap, Small, Low-Powered and Clickable Fixed-Icon Laser Projectors. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 389–394. https://doi.org/10.1145/2642918.2647356

[44] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. 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, 321–333. https://doi.org/10.1145/2984511.2984582

[45] Rana Abdulrahman Lateef and Ayad Rodhan Abbas. 2022. Human Activity Recognition using Smartwatch and Smartphone: A Review on Methods, Applications, and Challenges. *Iraqi Journal of Science* (2022), 363–379.

[46] Antonio Lazaro, Marc Lazaro, Ramon Villarino, and David Girbau. 2021. Seat-occupancy detection system and breathing rate monitoring based on a low-cost mm-wave radar at 60 ghz. *IEEE Access* 9 (2021), 115403–115414.

[47] Luis A. Leiva, Matjaz Kljun, Christian Sandor, and Klen Copic Pucihar. 2021. The Wearable Radar: Sensing Gestures Through Fabrics. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 17, 4 pages. https://doi.org/10.1145/3406324.3410720

[48] Ke Li, Ruidong Zhang, Bo Liang, François Guimbretière, and Cheng Zhang. 2022. EarIO: A Low-Power Acoustic Sensing Earable for Continuously Tracking Detailed Facial Movements. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 62 (jul 2022), 24 pages. https://doi.org/10.1145/3534621

[49] Yumeng Liang, Anfu Zhou, Huanhuan Zhang, Xinzhe Wen, and Huadong Ma. 2021. FG-LiquID: A Contact-less Fine-grained Liquid Identifier by Pushing the Limits of Millimeter-wave Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–27.

[50] Jaime Lien, Nicholas Gillian, M Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 1–19.

[51] Haipeng Liu, Yuheng Wang, Anfu Zhou, Hanyue He, Wei Wang, Kunpeng Wang, Peilin Pan, Yixuan Lu, Liang Liu, and Huadong Ma. 2020. Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 4, 4 (2020), 1–28.

[52] Haipeng Liu, Anfu Zhou, Zihe Dong, Yuyang Sun, Jiahe Zhang, Liang Liu, Huadong Ma, Jianhua Liu, and Ning Yang. 2022. M-Gesture: Person-Independent Real-Time In-Air Gesture Recognition Using Commodity Millimeter Wave Radar. *IEEE Internet of Things Journal* 9, 5 (2022), 3397–3415. https://doi.org/10.1109/JIOT.2021.3098338

[53] PL Lowbridge. 1995. Low cost millimeter-wave radar systems for intelligent vehicle cruise control applications. *Microwave Journal* 38, 10 (1995), 20–27.

[54] Guohua Lu, Fang Yang, Xijing Jing, and Jianqi Wang. 2010. Contact-free measurement of heartbeat signal via a doppler radar using adaptive filtering. In *2010 International Conference on Image Analysis and Signal Processing*. IEEE, 89–92.

[55] M. Marcus and B. Pattan. 2005. Millimeter wave propagation: spectrum management implications. *IEEE Microwave Magazine* 6, 2 (2005), 54–62. https://doi.org/10.1109/MMW.2005.1491267

[56] Tetiana Martyniuk, Orest Kupyn, Yana Kurlyak, Igor Krashenyi, Jiří Matas, and Viktoriia Sharmanska. 2022. Dad-3dheads: A large-scale dense, accurate and diverse dataset for 3d head alignment from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20942–20952.

[57] Denys JC Matthies, Alex Woodall, and Bodo Urban. 2021. Prototyping Smart Eyewear with Capacitive Sensing for Facial and Head Gesture Detection. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*. 476–480.

[58] Vimal Mollyn, Karan Ahuja, Dhruv Verma, Chris Harrison, and Mayank Goel. 2022. SAMoSA: Sensing Activities with Motion and Subsampled Audio. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–19.

[59] Louis-Philippe Morency and Trevor Darrell. 2006. Head Gesture Recognition in Intelligent Interfaces: The Role of Context in Improving Recognition. In *Proceedings of the 11th International Conference on Intelligent User Interfaces* (Sydney, Australia) *(IUI '06)*. Association for Computing Machinery, New York, NY, USA, 32–38. https://doi.org/10.1145/1111449.1111464

[60] Louis-Philippe Morency, Candace Sidner, Christopher Lee, and Trevor Darrell. 2005. Contextual recognition of head gestures. In *Proceedings of the 7th international conference on Multimodal interfaces*. 18–24.

[61] Nielsen Norman Group. 2022. Confirmation Dialogs Can Prevent User Errors — If Not Overused. https://www.nngroup.com/articles/confirmation-dialog/ Retrieved November 07, 2022.

[62] Amani Yousef Owda, Neil Salmon, Alexander J Casson, and Majdi Owda. 2020. The reflectance of human skin in the millimeter-wave band. *Sensors* 20, 5 (2020), 1480.

[63] Sameera Palipana, Dariush Salami, Luis A Leiva, and Stephan Sigg. 2021. Pantomime: Mid-air gesture recognition with sparse millimeter-wave radar point clouds. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–27.

[64] Keunwoo Park, Daehwa Kim, Seongkook Heo, and Geehyuk Lee. 2020. MagTouch: Robust Finger Identification for a Smartwatch Using a Magnet Ring and a Built-in Magnetometer. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376234

[65] Muneeba Raja, Zahra Vali, Sameera Palipana, David G. Michelson, and Stephan Sigg. 2020. 3D Head Motion Detection Using Millimeter-Wave Doppler Radar. *IEEE Access* 8 (2020), 32321–32331. https://doi.org/10.1109/ACCESS.2020.2973957

[66] Gabriel Reyes, Jason Wu, Nikita Juneja, Maxim Goldshtein, W. Keith Edwards, Gregory D. Abowd, and Thad Starner. 2018. SynchroWatch: One-Handed Synchronous Smartwatch Gestures Using Correlation and Magnetic Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 158 (jan 2018), 26 pages. https://doi.org/10.1145/3161162

[67] Gabriel Reyes, Dingtian Zhang, Sarthak Ghosh, Pratik Shah, Jason Wu, Aman Parnami, Bailey Bercik, Thad Starner, Gregory D. Abowd, and W. Keith Edwards. 2016. Whoosh: Non-Voice Acoustics for Low-Cost, Hands-Free, and Rapid Input on Smartwatches. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers* (Heidelberg, Germany) *(ISWC '16)*. Association for Computing Machinery, New York, NY, USA, 120–127. https://doi.org/10.1145/2971763.2971765

[68] Salah Saleh and Karsten Berns. 2015. Nonverbal Communication with a Humanoid Robot via Head Gestures. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '15)*. Association for Computing Machinery, New York, NY, USA, Article 15, 8 pages. https://doi.org/10.1145/2769493.2769543

[69] Farshid Salemi Parizi, Eric Whitmire, and Shwetak N. Patel. 2022. AuraRing: Precise Electromagnetic Finger Tracking. *GetMobile: Mobile Comp. and Comm.* 25, 3 (jan 2022), 34–37. https://doi.org/10.1145/3511285.3511295

[70] Samsung Developers. 2022. Dialogs. https://developer.samsung.com/one-ui-watch-tizen/ui-components/dialogs.html Retrieved November 07, 2022.

[71] Sirat Samyoun and John Stankovic. 2021. VoiSense: Harnessing Voice Interaction on a Smartwatch to Collect Sensor Data: Demo Abstract. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2021)*. 388–389.

[72] Panneer Selvam Santhalingam, Al Amin Hosain, Ding Zhang, Parth Pathak, Huzefa Rangwala, and Raja Kushalnagar. 2020. mmASL: Environment-independent ASL gesture recognition using 60 GHz millimeter-wave signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–30.

[73] Farhad Shahmohammadi, Anahita Hosseini, Christine E King, and Majid Sarrafzadeh. 2017. Smartwatch based activity recognition using active learning. In *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 321–329.

[74] Akash Deep Singh, Sandeep Singh Sandha, Luis Garcia, and Mani Srivastava. 2019. RadHAR: Human Activity Recognition from Point Clouds Generated through a Millimeter-Wave Radar. In *Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems* (Los Cabos, Mexico) *(mmNets'19)*. Association for Computing Machinery, New York, NY, USA, 51–56. https://doi.org/10.1145/3349624.3356768

[75] Texas Instruments. 2023. AWR1843 FMCW Radar Sensor. https://www.ti.com/lit/pdf/swrs222 Retrieved May 06, 2023.

[76] Texas Instruments. 2023. DCA1000EVM. https://www.ti.com/lit/pdf/spruij4 Retrieved May 06, 2023.

[77] Oleg Špakov and Päivi Majaranta. 2012. Enhanced Gaze Interaction Using Simple Head Gestures. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. Association for Computing Machinery, New York, NY, USA, 705–710. https://doi.org/10.1145/2370216.2370369

[78] Saiwen Wang, Jie Song, Jaime Lien, Ivan Poupyrev, and Otmar Hilliges. 2016. Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 851–860.

[79] Xue Wang and Yang Zhang. 2021. Nod to Auth: Fluent AR/VR Authentication with User Head-Neck Modeling. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 452, 7 pages. https://doi.org/10.1145/3411763.3451769

[80] Haowen Wei, Ziheng Li, Alexander D Galvan, Zhuoran Su, Xiao Zhang, Kaveh Pahlavan, and Erin T Solovey. 2022. IndexPen: Two-Finger Text Input with Millimeter-Wave Radar. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–39.

[81] Haolin Wei, Patricia Scanlon, Yingbo Li, David S Monaghan, and Noel E O'Connor. 2013. Real-time head nod and shake detection for continuous human affect recognition. In *2013 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*. IEEE, 1–4.

[82] Ting Wu, Theodore S Rappaport, and Christopher M Collins. 2015. The human body and millimeter-wave wireless communication systems: Interactions and implications. In *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2423–2429.

[83] Te-Yen Wu, Shutong Qi, Junchi Chen, MuJie Shang, Jun Gong, Teddy Seyed, and Xing-Dong Yang. 2020. Fabriccio: Touchless gestural input on interactive fabrics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[84] Haijun Xia, Tovi Grossman, and George Fitzmaurice. 2015. NanoStylus: Enhancing Input on Ultra-Small Displays with a Finger-Mounted Stylus. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software amp; Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 447–456. https://doi.org/10.1145/2807442.2807500

[85] Yucheng Xie, Ruizhe Jiang, Xiaonan Guo, Yan Wang, Jerry Cheng, and Yingying Chen. 2022. Universal Targeted Attacks against MmWave-Based Human Activity Recognition System. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services* (Portland, Oregon) *(MobiSys '22)*. Association for Computing Machinery, New York, NY, USA, 541–542. https://doi.org/10.1145/3498361.3538774

[86] Zihan Yan, Yufei Wu, Yiyang Li, Yifei Shan, Xiangdong Li, and Preben Hansen. 2022. Design Eye-Tracking Augmented Reality Headset to Reduce Cognitive Load in Repetitive Parcel Scanning Task. *IEEE Transactions on Human-Machine Systems* 52, 4 (2022), 578–590.

[87] Zihan Yan, Yue Wu, Yifei Shan, Wenqian Chen, and Xiangdong Li. 2022. A dataset of eye gaze images for calibration-free eye tracking augmented reality headset. *Scientific Data* 9, 1 (2022), 1–16.

[88] Zihan Yan, Yufei Wu, Yang Zhang, and Xiang 'Anthony' Chen. 2022. EmoGlass: An End-to-End AI-Enabled Wearable Platform for Enhancing Self-Awareness of Emotional Health. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 13, 19 pages. https://doi.org/10.1145/3491102.3501925

[89] Xiaoying Yang and Yang Zhang. 2021. CubeSense: Wireless, Battery-Free Interactivity through Low-Cost Corner Reflector Mechanisms. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 386, 6 pages. https://doi.org/10.1145/3411763.3451599

[90] Zi-Kai Yang, Heping Shi, Sheng Zhao, and Xiang-Dong Huang. 2020. Vital sign detection during large-scale and fast body movements based on an adaptive noise cancellation algorithm using a single Doppler radar sensor. *Sensors* 20, 15 (2020), 4183.

[91] Hui-Shyong Yeo, Gergely Flamich, Patrick Schrempf, David Harris-Birtill, and Aaron Quigley. 2016. RadarCat: Radar Categorization for Input amp; Interaction. 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, 833–841. https://doi.org/10.1145/2984511.2984515

[92] Hui-Shyong Yeo, Juyoung Lee, Andrea Bianchi, and Aaron Quigley. 2016. WatchMI: Pressure Touch, Twist and Pan Gesture Input on Unmodified Smartwatches. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 394–399. https://doi.org/10.1145/2935334.2935375

[93] Shanhe Yi, Zhengrui Qin, Ed Novak, Yafeng Yin, and Qun Li. 2016. Glassgesture: Exploring head gesture interface of smart glasses. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 1–9.

[94] Maxim Zhadobov, Nacer Chahat, Ronan Sauleau, Catherine Le Quement, and Yves Le Drean. 2011. Millimeter-wave interactions with the human body: state of knowledge and recent advances. *International Journal of Microwave and Wireless Technologies* 3, 2 (2011), 237–247. https://doi.org/10.1017/S1759078711000122

[95] Cheng Zhang, AbdelKareem Bedri, Gabriel Reyes, Bailey Bercik, Omer T. Inan, Thad E. Starner, and Gregory D. Abowd. 2016. TapSkin: Recognizing On-Skin Input for Smartwatches. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces* (Niagara Falls, Ontario, Canada) *(ISS '16)*. Association for Computing Machinery, New York, NY, USA, 13–22. https://doi.org/10.1145/2992154.2992187

[96] Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Sumeet Jain, Yiming Pu, Sinan Hersek, Kent Lyons, Kenneth A. Cunefare, Omer T. Inan, and Gregory D. Abowd. 2017. SoundTrak: Continuous 3D Tracking of a Finger Using Active Acoustics. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 30 (jun 2017), 25 pages. https://doi.org/10.1145/3090095

[97] Fusang Zhang, Jie Xiong, Zhaoxin Chang, Junqi Ma, and Daqing Zhang. 2022. Mobi2Sense: empowering wireless sensing with mobility. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*. 268–281.

[98] Jia Zhang, Yinian Zhou, Rui Xi, Shuai Li, Junchen Guo, and Yuan He. 2022. AmbiEar: MmWave Based Voice Recognition in NLoS Scenarios. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 151 (sep 2022), 25 pages. https://doi.org/10.1145/3550320

[99] Ruidong Zhang, Mingyang Chen, Benjamin Steeper, Yaxuan Li, Zihan Yan, Yizhuo Chen, Songyun Tao, Tuochao Chen, Hyunchul Lim, and Cheng Zhang. 2022. SpeeChin: A Smart Necklace for Silent Speech Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 192 (dec 2022), 23 pages. https://doi.org/10.1145/3494987

[100] Yang Zhang, Junhan Zhou, Gierad Laput, and Chris Harrison. 2016. SkinTrack: Using the Body as an Electrical Waveguide for Continuous Finger Tracking on the Skin. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1491–1503. https://doi.org/10.1145/2858036.2858082

[101] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. 2018. Through-wall human pose estimation using radio signals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

7356–7365.

[102] Peijun Zhao, Chris Xiaoxuan Lu, Bing Wang, Niki Trigoni, and Andrew Markham. 2023. CubeLearn: End-to-end learning for human motion recognition from raw mmWave radar signals. *IEEE Internet of Things Journal* (2023).

[103] Peijun Zhao, Chris Xiaoxuan Lu, Jianan Wang, Changhao Chen, Wei Wang, Niki Trigoni, and Andrew Markham. 2019. mid: Tracking and identifying people with millimeter wave radar. In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 33–40.

[104] Gang Zhou. 2015. Automobile Anti-collision Millimeter-Wave Radar Signal Processing. In *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, Vol. 2. IEEE, 484–486.

[105] Junhan Zhou, Yang Zhang, Gierad Laput, and Chris Harrison. 2016. AuraSense: Enabling Expressive Around-Smartwatch Interactions with Electric Field Sensing. 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, 81–86. https://doi.org/10.1145/2984511.2984568