

Exploring Uni-manual Around Ear Off-Device Gestures for Earables

SHAIKH SHAWON AREFIN SHIMON, School of Computer Science, University of Waterloo, Canada

ALI NESHATI, Ontario Tech University, Canada

JUNWEI SUN and QIANG XU, Huawei Human-Machine Interaction Lab, Canada

JIAN ZHAO, School of Computer Science, University of Waterloo, Canada

Small form factor limits physical input space in earable (i.e., ear-mounted wearable) devices. Off-device earable inputs in alternate mid-air and on-skin around-ear interaction spaces using uni-manual gestures can address this input space limitation. Segmenting these alternate interaction spaces to create multiple gesture regions for reusing off-device gestures can expand earable input vocabulary by a large margin. Although prior earable interaction research has explored off-device gesture preferences and recognition techniques in such interaction spaces, supporting gesture reuse over multiple gesture regions needs further exploration. We collected and analyzed 7560 uni-manual gesture motion data from 18 participants to explore earable gesture reuse by segmentation of on-skin and mid-air spaces around the ear. Our results show that gesture performance degrades significantly beyond 3 mid-air and 5 on-skin around-ear gesture regions for different uni-manual gesture classes (e.g., *swipe*, *pinch*, *tap*). We also present qualitative findings on most and least preferred regions (and associated boundaries) by end-users for different uni-manual gesture shapes across both interaction spaces for earable devices. Our results complement earlier elicitation studies and interaction technologies for earables to help expand the gestural input vocabulary and potentially drive future commercialization of such devices.

CCS Concepts: • **Computing methodologies** → *Motion capture*; • **Human-centered computing** → **Mobile devices**; **Empirical studies in ubiquitous and mobile computing**; **Gestural input**; • **Hardware** → **Emerging interfaces**.

Additional Key Words and Phrases: Embodied Interaction, Input Techniques, Uni-manual Interaction, Touch Surfaces, Ear-based Interaction, Earables.

ACM Reference Format:

Shaikh Shawon Arefin Shimon, Ali Neshati, Junwei Sun, Qiang Xu, and Jian Zhao. 2024. Exploring Uni-manual Around Ear Off-Device Gestures for Earables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 1, Article 3 (March 2024), 29 pages. <https://doi.org/10.1145/3643513>

1 INTRODUCTION

Wireless earbuds are one of the most popular wearables [66] among consumers. It is now commonplace to see people wearing such wearables in public settings (e.g., office, gym, public transportation) to discretely listen to music, take calls, or filter out external noise using noise suppression features for creating personal quiet zones. However, the small form factor and input area constraint in wireless earbuds restrict physical touch interaction to simple on-device taps and swipes, thus limiting their input space. Off-the-shelf wireless earbuds increasingly integrate different sensors [55] to extend their functionality beyond personal audio and hearing aid support [11, 46, 49]. Wireless earbuds such as *Apple AirPods* [4], or *Samsung Galaxy Buds Pro* [53] are pre-equipped with IMU, motion, and health sensors for supporting health sensing and fitness tracking. These kinds

Authors' addresses: Shaikh Shawon Arefin Shimon, ssarefin@uwaterloo.ca, School of Computer Science, University of Waterloo, 200 University Ave W, Waterloo, Ontario, Canada, N2L 3G1; Ali Neshati, ali.neshati@ontariotechu.ca, Ontario Tech University, 2000 Simcoe St N, Oshawa, Ontario, Canada, L1G 0C5; Junwei Sun, junwei.sun@huawei.com; Qiang Xu, qiang.xu1@huawei.com, Huawei Human-Machine Interaction Lab, 19 Allstate Pkwy, Markham, Ontario, Canada; Jian Zhao, jianzhao@uwaterloo.ca, School of Computer Science, University of Waterloo, 200 University Ave W, Waterloo, Ontario, Canada, N2L 3G1.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, <https://doi.org/10.1145/3643513>.

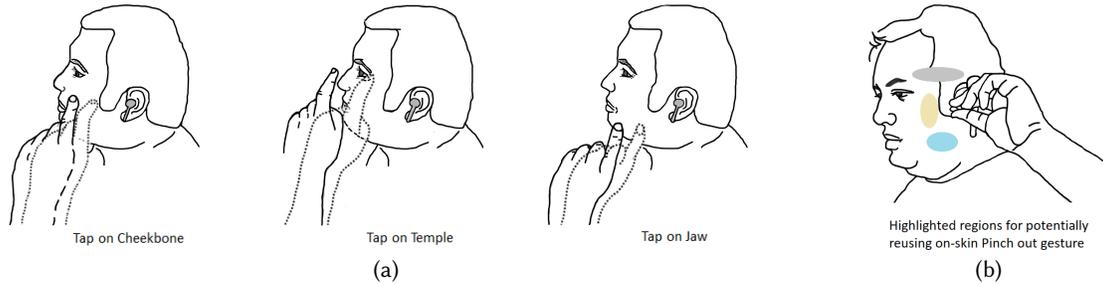


Fig. 1. Reusing gestures at different regions for off-device Earable interaction: (a) Off-device, around-ear and on-skin taps at 3 distinct gesture regions. (b) Potential gesture regions with highlighted boundaries.

of sensor-infused, ear-mounted computing devices supporting additional functionalities (e.g., health sensing [44], activity recognition [24], gesture sensing [41], wearable public display [58]) beyond traditional personal audio or hearing aids are defined as *earable devices* [51] in the ubiquitous computing community.

Earable platforms have gained significant traction among ubiquitous computing researchers [11, 30, 46], especially in supporting gestural sensing to counter the on-device interaction space limitation problem. The research community has proposed several off-device approaches [51] (e.g., manual on-skin/mid-air motion gestures, head movements, facial expressions, voiced and silent commands) leveraging gesture sensing using off-the-shelf or custom hardware to expand the earable input vocabulary. Among the proposed approaches, mid-air and on-skin uni-manual off-device around-ear gestures are popular strategies [8, 11]. Such hand-to-face gestures tend to occur naturally and unconsciously with significant daily frequency among end users [34, 45]. Besides gesture recognition techniques for such off-device gestures [51] for earables, prior gesture elicitation studies [9, 47] looked at eliciting gesture vocabulary with individual qualitative metrics such as *task suitability*, *gesture usability*, and *social comfort*. These studies revealed that end-users have a high agreement regarding uni-manual horizontal (front/back) and vertical (up/down) *swipe*, *pinch (in/out)*, and *tap* gestures for off-device earable interactions.

Such on-skin and mid-air gestures could be fine-grained by reusing them across different gesture regions in an interaction space (Figure 1a). For example, a *tap* on the temple could toggle the *noise cancellation* feature on an earable device. In contrast, a *tap* on the cheek could toggle the *microphone muting* feature on the earable device. Understanding the threshold for the number of unique gesture regions in an interaction space for different gestures and their associated region boundaries is necessary to support gesture reuse in off-device and around-ear earable interaction. Segmenting an interaction space into multiple gesture regions could greatly expand the gestural input vocabulary while reusing a few gesture classes (*swipe*, *tap*, *pinch*), improving gesture memorability [43]. Although prior literature explored off-device uni-manual gesture reuse across different segmented regions for outer-ear touches to support gesture reusability, similar exploration is lacking for around-ear on-skin space (i.e., face, head, and neck) and above-ear mid-air spaces for supporting similar gestures.

We address this research gap by studying the reuse of different gesture classes at different mid-air and on-skin gesture regions and the effect of increasing the number of gesture regions on quantitative and qualitative gesture performance metrics for mid-air and on-skin interaction spaces. By leveraging some common uni-manual off-device and around-ear gestures (i.e., *swipe*, *tap*, and *pinch*) proposed in earlier earable gesture elicitation studies [38, 47], we collected 7,560 gesture motion data from 18 participants using a *Vicon* motion camera for analyzing their performance against an increasing number of gesture regions in mid-air and on-skin interaction spaces. We also examined how participants defined gesture region boundaries in the off-device interaction spaces under consideration.

The following outlines the contributions of our work:

- Comparison of mid-air and around-ear on-skin space for uni-manual, off-device earable interaction across different gesture classes;
- Understanding of end-user preference shift from one interaction space to another for increasing gesture reuse;
- Analysis of the effects of different factors (i.e., the number of gesture regions and associated boundaries) on interaction space segmentation for uni-manual, off-device earable gestures.

Collectively, the findings from our gesture motion experiment on end-users for off-device, around ear earable gestures provide a guideline for implementing gesture reuse across different interaction spaces by complimenting earlier gesture recognition technology research for earables and qualitative exploration of gesture preferences for off-device earable interaction.

2 RELATED WORK

2.1 Gesture Recognition Support and Device Format for Earables

Prior research on interaction technologies for alternate earable interaction spaces looked at face touch [29, 62], ear touch [31, 36, 65], mid-air manual motion gesture [2, 25, 26, 28, 37, 39, 40, 59], head movements [17, 60], facial expressions [3, 10, 35, 56], voice [6], and silent [27, 54, 57, 64] command recognition to solve earable input space limitation by leveraging various sensors (IMU, proximity, infrared, PPG, ECG) and microphones present in off-the-shelf, or custom earable hardware. Besides these sensor-based approaches, computer vision has also been utilized in various instances [26, 37, 39, 59] to detect various mid-air or hand-to-face gestures using cameras mounted on earable devices. However, such camera-based, body-mounted gesture recognition platforms have practical usability and privacy concerns [27]. Although research into end-user acceptability for gestures shows that facial expression-based gestures are acceptable in private settings, they are less socially acceptable in public settings [32] due to the possibility of attracting unwarranted, negative attention [13]. Voice or silent command-based interactions require memorization of a large set of unique commands, and are not well suited for user interface navigational tasks [16] in other wearable or interactive devices. Physical uni-manual interactions on ear helix (e.g., ear bends, pinch, slides) proposed in Lissermann et al. [36] can potentially displace earable devices, making on-ear physical touch-based interaction difficult for expanding earable input space. Head/neck-based movement can put significant physical strain after repetitive use. Large manual gestures away from body suffer from “gorilla arm” fatigue syndrome [22] and can attract unwarranted, negative attention [13] in public spaces, similar to facial and head/neck motion gestures.

In natural, unconstrained settings when one or both hands are free for gestural interaction, studies have shown that end-users prefer making uni-manual hand-to-face gestures on skin or mid-air closer to the head, compared to other large body gestures [23]. Such hand-to-face gestures are unconscious and spontaneous motions [5, 14, 21, 33, 42] with observed frequency from 15.7 to 23 contacts/hr [45]. Such subtle hand-to-face gestures are more socially acceptable [32] than large manual motion gestures. Previous research on earables [36, 47] points out some real-world scenarios of uni-manual interaction with ear-based devices in real-world settings. Such interaction also closely matches end-user expectations of wearable earbud devices, where user behavior is to primarily interact with one earbud with one hand at a time in a public space, freeing up the alternate hand for other possible tasks.

These reasons make uni-manual touch gestures on the skin around the ear (face, head, and neck) and mid-air gestures around the head viable choices for expanding earable input vocabulary. Most of the previously discussed research on on-skin and mid-air earable interaction spaces do show that it is possible to recognize both gesture classes and gesture location over different areas around the face. However, the prior literature lacks a comprehensive analysis of earable gesture spaces to support gesture reuse by leveraging gesture location and gesture class recognition techniques. This inspired our research focus on off-device gesture space exploration to

support gesture reuse in earables and identify gesture regions of interest. Research into emotional user experience elements of earphone form factors [63] shows in-ear earables (i.e., wireless earbuds) have higher wearing style satisfaction and preference for discrete interaction with device in public spaces [48] compared to behind-the-ear or on-ear (e.g., headphones) earable formats. As such, we limited our gesture reuse study to in-ear earable devices such as wireless earbuds.

2.2 Interaction Space Segmentation for Gestural Input Devices

Wong et al. [61] explored *bezel-initiated swipe* (BIS—a form of uni-manual touch interaction) reuse over multiple gesture regions (6/ 8/ 12/ 16/ 24/ 36) on smartwatch bezels. *Accuracy*, *task completion time*, and other round-bezel-specific metrics (i.e., *absolute/relative offset*) were used to measure BIS performance against increasing gesture regions for input reuse. Rey et al. [50] proposed a similar uni-manual smartwatch gesture called *bezel-to-bezel* (B2B) interaction and explored the reuse of the B2B gesture over 4, 6, and 8 bezel regions. Dezfuli et al. [12] explored segmentation of an imaginary gesture space over hand/palm surface for using on-skin *swipe* and *tap*-like gestures to control a smart TV. Their observations revealed that up to 5 landmarks on the palm could be correctly targeted using uni-manual touch gestures with alternate hands. Gil et al. [19] proposed an in-air thumb typing system (*ThumbAir*) on a commercial head-mounted display (HMD), where they explored viable gesture region positions and an optimal number of regions for dividing the input space for different letter input groups. Their analysis leveraged similar general metrics to [50, 61] (e.g., *accuracy*, *gesture time*) and gesture motion-specific metrics (e.g., *wrist angle change*) to compare the different numbers of regions on both hands for typing input. Although the above works focus on different types of wearable devices compared to earables, the analysis metrics and study design still provide essential guidelines for performing analysis of wearable gesture reuses over multiple gesture regions in alternate interaction spaces.

Besides proposing custom earable hardware to detect physical touch gestures on the outer ear, Lissermann et al. [36] explored the ear helix design space for touch gesture reuse across multiple regions, examined the relationship between gesture accuracy and increasing gesture regions for gesture reuse, and suggested potential applications leveraging this design space for earable interaction. However, their exploration into region segmentation for earable interaction is limited only to the outer ear and explored in the context of a single touch or tap gesture. The effect of other touch primitives (e.g., swipes and pinches) on the outer ear has not been explored in prior works. Prior earable research also lacks exploration of around-ear on-skin/mid-air gesture reuse.

To address this research gap, we focus on exploring the segmentation of on-skin (i.e., face/neck or other parts of the head) and mid-air (i.e., above-ear and around-head), around-ear interaction space to support gesture reuse across different gesture regions. Our work is closely related to and inspired by gesture region segmentation analysis on ear helix for uni-manual physical interaction in *Earput* [36]. Based on findings from previous earable elicitation studies [9, 47], we limit our gesture classes to primitive gestures such as *tap*, *swipe*, and *pinch* for exploring gesture reusability across multiple gesture regions.

3 METHODOLOGY

3.1 Research Questions

In this work, we experimented with 7 gestures (*swipe up*, *swipe down*, *swipe front*, *swipe back*, *pinch in*, *pinch out*, and *tap*). These were the most common gestures proposed in the previous earable elicitation studies [9, 47] and had the most ease of performance, social acceptability, and memorability in various settings [1]. We did not leverage the “*cover ear*” gesture elicited from [47] study as our study focused only on around-ear gestures. We explored how users perform the gestures mentioned above within the following two interaction spaces:

- **On-skin, around-ear space:** This space includes all areas on the skin above the neck, excluding both ears.

- **Mid-air space:** This space includes mid-air space above and around the ear without any limitation of distance from the ear.

We aimed to analyze the gesture motion data from the 7 gestures as mentioned earlier and identify region boundaries for gesture reuse in on-skin and mid-air interaction spaces to answer the following research questions:

- **RQ1:** Whether and how does gesture performance vary between in mid-air and on-skin space?
- **RQ2:** Whether and how does gesture performance vary when the number of segments increases in a chosen interaction space?
- **RQ3:** Whether and how does end-user consensus exist on most and least preferred regions across a fixed number of gesture regions in a particular interaction space?

3.2 Experimental Variables and Hypotheses

To answer our research questions, we designed an experiment using the following 3 independent variables:

- **IV1 - Interaction space:** On-skin and Mid-air.
- **IV2 - Gesture shape:** Tap, Swipe Up, Swipe Down, Swipe Front, Swipe Back, Pinch In, and Pinch Out.
- **IV3 - Number of gesture regions:** 3, 5, and 7.

The following dependent variables were leveraged to measure the performance of gestures across different interaction spaces and among different numbers of gesture regions in a particular interaction space.

- **DV1 - Gesture time:** Time window measured from the end of a *gesture delimiter* to the end of the gesture recording, during which the user performs a given gesture G , measured in milliseconds (ms). This metric was chosen based on prior related interaction space segmentation research [50, 61]. Details on calculating DV1 with the help of delimiter gestures are discussed in Section 4.2.
- **DV2 - Gesture path length:** Length of gesture G traversed during the above (DV1) time window, measured in millimeters (mm). This metric was developed specifically for this study, as we wanted to analyze the starting and ending of a gesture across different regions in an interaction space.
- **DV3 - Gesture accuracy:** The percentage of the actual number of times (n) a gesture G was successfully done on (on-skin) / above (mid-air) target gesture regions in an interaction space (S) with respect to the required number of trials. This metric was chosen based on prior research [50, 61].

These quantitative criteria allow us to compare the performance of gestures in each interaction space and observe the effect of the number of segments in the interaction space on gesture performance, thus helping us investigate **RQ1** and **RQ2**. We formulated the following two null hypotheses based on our research questions.

- H_0^{RQ1} : User performance (DV1, DV2, DV3) does not vary when gesture G (IV2) is performed in different interaction spaces (IV1).
- H_0^{RQ2} : User performance (DV1, DV2, DV3) for gesture G (IV2) does not vary when the number of segments (IV3) increase for a specific interaction space (IV1).

In addition, to investigate **RQ3**, we aimed to employ qualitative observations based on NASA TLX surveys, collected gesture region definitions, questionnaire results, and participant discussions.

4 EXPERIMENT

4.1 Participants

We recruited 18 participants (11 males and 7 females) aged between 22 and 40 ($M = 26.37$, $SD = 5.520$). All the participants were right-handed and wore a watch in their left hand. Participants reported varying levels of earbud usage in their regular life. Overall, 17 out of 18 participants reported using wireless earbuds to make calls, listen to music, and participate in online meetings using phones and workstations (period: 0–25 hours/week,



Fig. 2. Experimental setup with Vicon: the participant is performing gestural tasks shown on the screen

$M = 9.1$, $SD = 7.12$). The study was divided into two halves corresponding to two interaction spaces with a break in between. Each trial lasted 15-25 seconds, and each half of the study took 75 to 90 minutes on average. The total study took around 2.5 to 3 hours per participant, with pre-study and post-study times. Each participant was paid at a rate of \$15 per hour for their time and effort.

4.2 Apparatus

We designed a desktop application to record, filter, and analyze the gesture data, henceforth called *Gesture Recording Data Collection Application* (GRDA). GRDA was configured on a desktop PC connected to 8 *Vicon* cameras and a 24-inch screen (Figure 2). This PC served as the *Vicon* camera server and was configured with an *Intel Xeon* processor with 32 GB RAM and 16 GB graphics card attached with a standard keyboard and mouse for running *Vicon Tracker 3.9* application.

The gesture recording apparatus was set up to support both right and left-handed participants. As all participants in our study were right-handed, they were asked to create uni-manual gestures using their non-dominant (left) hand so that the dominant (right) hand could control GRDA using a keyboard and mouse for data recording. The *Vicon* cameras recorded the gestural inputs using markers attached to the participant's head and left hand¹. For gesture tracking and reference purposes, 3 markers were attached to different parts of the head using velcro straps (Figure 2). The markers were positioned in the following places: 1) center top on the crown of the head; 2) back of the head; and 3) side of the head behind the temple and above the helix of the left ear². Finger markers were attached to the left index (and thumb) finger¹ using double-sided tape. One marker was attached to the index finger for all tasks, as shown in Figure 2. For specific gestural inputs (pinch in/pinch out) involving two fingers, another marker was also placed on the left¹ thumb finger. Before running GRDA for our experiment, the *Vicon Tracker 3.9* application was turned on to track the gesture motion data using the *Vicon* markers and continuously send the data stream back to GRDA for recording purposes. The participant controlled gesture data recording using this GRDA application (Figure 2), which filtered and categorized the marker data stream collected from the *Vicon Tracker 3.9* application running on the backend. All uni-manual gestures started from a fixed position (marked by a white tape in the table containing the *Vicon* server PC in the rightmost image of Figure 2). Before any actual gesture, participants performed one of the following two *Delimiter Gestures* above the table's fixed (taped) position.

¹ Non-dominant hand used for creating unimanual gestures. For left-handed participants, markers would be worn on the right hand.

² Closest ear to the hand used for creating uni-manual gesture. For left-handed participants, markers would be worn on the right hand and the head marker above the ear helix would be placed above the right ear.

- *Delimiter Gesture A*: Wiggle the index finger (3 times) with the index and thumb fingers of the marker-wearing left¹ hand. Used only before two-finger uni-manual gestures (pinch in/out) using index and thumb fingers.
- *Delimiter Gesture B*: Wiggle with the marker-wearing hand's index finger (3 times). Used before all single finger-based uni-manual gestures (tap, horizontal/vertical swipe) and during *head model generation* phase (Section 4.4.1).

The purpose of these delimiter gestures was to help: 1) accurately distinguish between the thumb and index finger of the marker-wearing hand for two-finger gestures and 2) mark the starting of the gesture for both the participant and the gesture apparatus. After the delimiter gesture was made and the motion for the targeted uni-manual gesture began, gesture metric measurement automatically started once the marker-wearing fingers reached within 30 cm from the center³ of the head (Figure 15 in Appendix A). The metric measurement stopped once the gesture was completed, and the hand moved out of the above range to filter out gesture motion noise and evaluate only the relevant gesture motion. Gesture recording automatically stopped when the participant completed the task and rested his hand on the starting position.

4.3 Design

We employed a within-subjects design for our experiment. According to the independent variables (IV1-3), each participant was provided with 2 interaction spaces (on-skin and mid-air) \times 3 segment numbers (3, 5, and 7) \times 7 gestures (*Tap*, *Swipe Up*, *Swipe Down*, *Swipe Front*, *Swipe Back*, *Pinch In*, and *Pinch Out*) = 42 experimental conditions. For each condition, the participant was requested to define the region boundaries on a paper containing a side view and a front view of a face, and the regions were then labeled with identifiers. Using a *Latin square* design, we counter-balanced the interaction space (IV1) and the number of segments (IV3) variables across the whole experiment. The participant was first given a particular interaction space, within which the participant was provided with a specific number of segments. For each combination of these two variables, the participant was asked to perform each of the seven gestures (IV2) in a randomized order, and there were 2 repetitions for each gesture task. The participant could perform different numbers of trials based on the number of segments within the IV1 and IV3 combination. For example, with mid-air and 5 segments, the participant would be asked to perform each of the seven gestures for 5 segments \times 2 repetitions = 10 times. The 5 segments labeled using region identifiers (A–E) were shown randomly across the tasks. Therefore, each participant performed: 2 interaction spaces \times (3 + 5 + 7) segments \times 7 gestures \times 2 repetitions = 420 trials. We collected 7,560 gestural data points for 18 participants.

4.4 Procedure

The experiment was conducted in an office environment with adequate soundproofing to provide a calm environment. During the experiment, the participants sat in front of a table containing the *Vicon* server PC and designated starting position (marked by white tape). They operated GRDA via keyboard and mouse to record the gesture, as in Figure 2. Before the study started, the purpose of the research was explained to the participants. They were encouraged to ask the researcher questions for clarification regarding any steps of the study. Participants were informed that they could take breaks or abort the experiment at any time. After explaining the study purpose and break structures, participant consent was obtained using a consent form. Then, participants began the two-phased study (described in Sections 4.4.1 and 4.4.2). During both phases, all gesture motions started from the table's designated position (marked with white tape). At the end of the study, demographic data was collected from the participants using a *Qualtrics* form.

³For the purpose of our calculation, the center of the head is located on the vertical plane bisecting the body, located between the eyes and the top of the tragus in the outer ear on the horizontal field-of-view (FOV) plane.

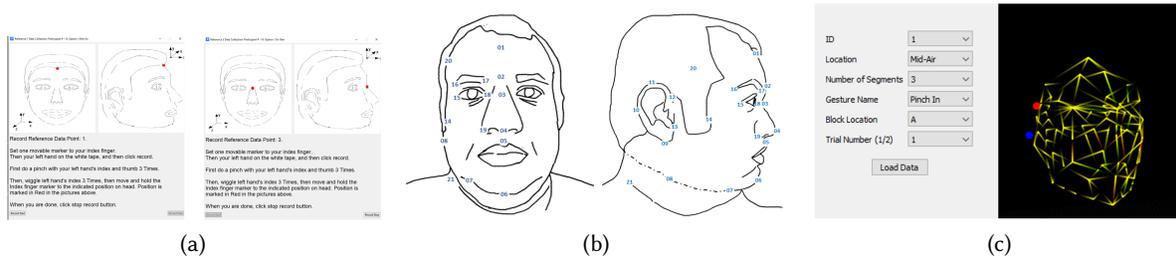


Fig. 3. (a) Collection of original reference points on the face using GRDA for head reconstruction. (b) The location of 21 original reference points on the face. (c) Snapshot of GRDA with a reconstructed head model.

4.4.1 Phase 1: Head Model Generation. As each participant has a unique facial structure, we first generated a 3D head model for each participant using the *Vicon* motion tracking camera and markers to collect meaningful gesture data in Phase 2. Figure 3c shows a sample of the reconstructed head model with GRDA. The head model was created by collecting 21 reference points on different head parts (e.g., *chin, earlobe, cheek, top of nose, temple*) and performing data augmentation.

For collecting each reference point data, GRDA showed the position of the reference point for the face, head, and neck. Once a referent point was shown (Figure 3A), the gesture recording task was explained in the application, with the researcher providing additional clarifications. Once participants were ready for reference data collection, they were asked to reposition their left ¹ hand at the designated starting point on the table. Once the hand is positioned, participants were instructed to start the recording via GRDA using the available dominant (right) hand, then perform the delimiter gesture using the marker wearing left ¹ hand, followed by holding the left index finger ⁴ on the designated reference point on the head for 3 seconds before stopping the recording. After the recording was completed, a pop-up recording confirmation window allowed the participants to proceed to the following reference point recording or redo the current recording if necessary. After all reference points were collected, the generated head model was shown on GRDA, allowing participants to adjust the head model using a custom toolkit integrated into GRDA. Once the head model was finalized, participants proceeded to Phase 2 to perform the actual gesture data collection task.

4.4.2 Phase 2: Gesture Data Collection. After confirmation of a head model, participants were shown 42 gesture conditions in an approach described in the study design (Section 4.3). For each condition, they were asked to define preferred region boundaries for the targeted interaction space and the number of around-ear gesture regions. They were also informed that region boundaries for different gesture conditions could vary. The participants were allowed to explore targeted interaction space in a practice session for each condition to build a mental model of gesture space segmentation and finalize the gesture region boundaries. For on-skin gestures, participants were reminded that touching the outer ear was prohibited, and touching any other region above the neck to the head crown was allowed. During mid-air gesture sessions, they were informed that physical touch interaction with the body part was prohibited. However, participants were clarified that mid-air regions could be directly above the outer ear and located near or away from the body, based on their preferences.

After completing a practice session, a side and front view image of the head and neck region on paper was presented to the participants to define the gesture region boundaries. Each region was marked alphabetically, starting with the letter **A**. The researcher asked specific questions about the positioning of each region to different body parts (i.e., eyebrow, top of nose, cheekbone) and took notes while the participant defined the region boundaries. Once boundaries were defined, a particular gesture region identifier (**A-G**) was presented to the

⁴For a right-handed participant, the left index finger contains a *Vicon* marker, which switches to right index finger for a left-handed participant.

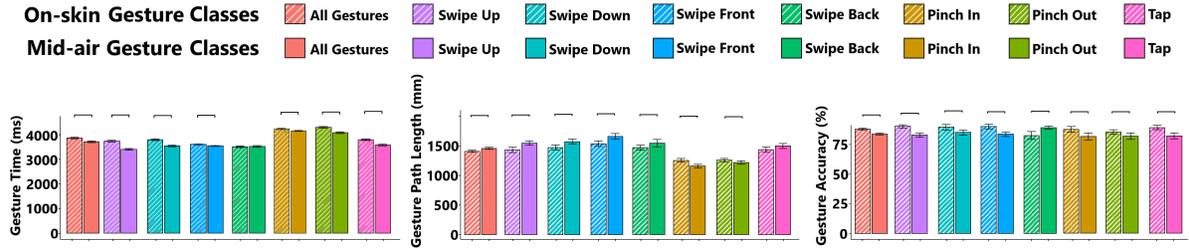


Fig. 4. Performance metrics for mid-air and on-skin gestures with 95% confidence intervals. Number of regions around the ears have been ignored.

participants on the desktop application (Figure 2) for recording uni-manual gestures for that region. After starting the gesture recording, participants began uni-manual motion from the starting position with the appropriate delimiter gesture, followed by the target gesture over the intended boundary before turning off the recording. Controlling the recording via GRDA was done using the right (dominant) hand, and simultaneously, the other hand was used for uni-manual gestures. After performing all the gesture trials for a specific gesture condition (interaction space and associated number of regions), a *Qualtrics* survey was presented in GRDA to rate the condition using a NASA TLX survey and identify preferred gesture region locations for that specific condition. After completing all conditions, an end-of-study *Qualtrics* survey was used to collect additional information regarding gesture preferences for both interaction spaces from participants.

5 QUANTITATIVE RESULTS

Using normality tests, we tested the data distribution of gesture metrics DV1, DV2, and DV3. When applicable, we followed up with testing homogeneity of variance for the distributions using Bartlett’s test for choosing appropriate parametric or non-parametric tests. Appendix B provides the supporting statistical analysis accompanying quantitative results outlined in Figures 4 and 5. Appendix B.1 outlines the mean and median gesture metric values across interaction spaces and the number of regions. Appendix B.2, B.3, and B.4 outline normality and significance test results to test the effect of interaction spaces (RQ1) and increasing around-ear regions (RQ2).

As the layouts and size for mid-air and on-skin regions associated with different gestures could vary between participants, *gesture time* (DV1) and *path length* (DV2) for different gestures could not be paired across interaction spaces for addressing RQ1 or matched across different numbers of around-ear gesture regions within each interaction space for addressing RQ2. However, as *gesture accuracy* (DV3) was not directly related to gesture region layout/size at different around-ear interaction spaces, this metric could be paired/matched for both research questions. These paired (on-skin and mid-air factors) or matched (3/5/7 regions) criteria guided our choice between paired/matched parametric and non-parametric tests for RQ1 and RQ2.

5.1 Effect of Interaction Space Choice on Performance Metrics (RQ1)

When analyzing *gesture time* (DV1) for individual gesture classes, we observed that overall (Figure 4), all gestures except *swipe back* took significantly more time to perform on-skin than in mid-air in general (Table 4 in Appendix B.2). Diving deeper (Figure 5), we observed that on-skin *swipe back* gestures took more time than mid-air counterparts across all (3/5/7) numbers of regions around the ear. However, the difference was statistically insignificant (Table 4). For all other gestures, we observed that on-skin gesture time was significantly higher than mid-air gestures for 3 and 5 around-ear regions. For 7 around-ear regions, only vertical (up/down) *swipe*, *pinch* (in/out), and *tap* gestures took significantly longer to perform on the skin. Although on-skin gesture time for horizontal (front/back) *swipe* gestures in 7 regions was higher, it was not statistically significant.

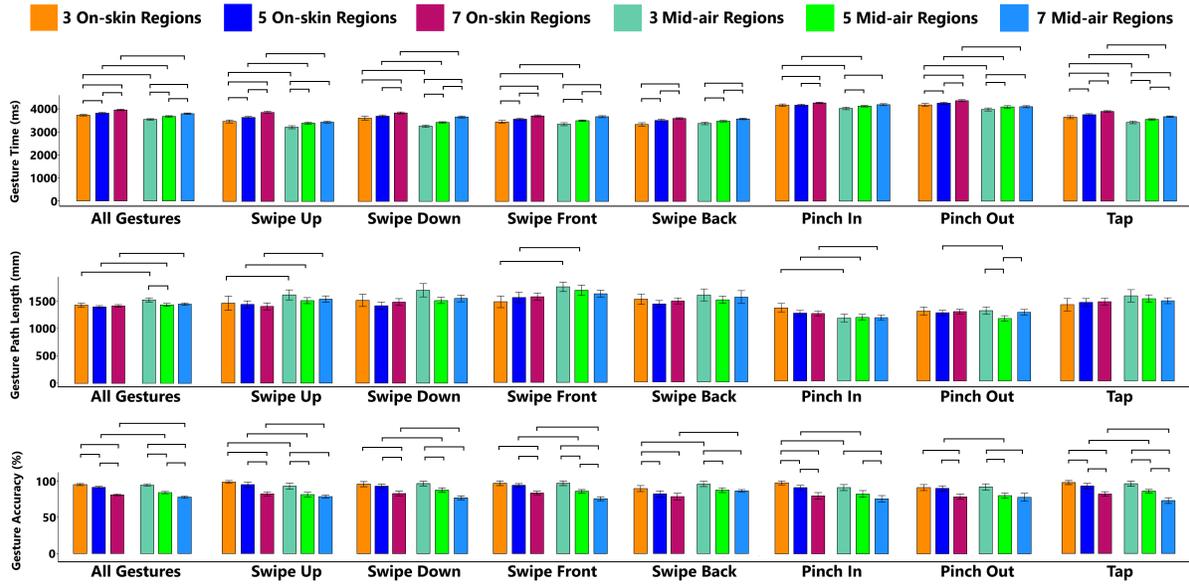


Fig. 5. Performance metrics for mid-air and on-skin gestures across all numbers of gesture regions around the ear with 95% confidence intervals.

Our *path length* (DV2) analysis also showed a significant effect of interaction space choice on gestures (Figure 4 - all gestures). In general, mid-air swipe and tap gesture paths were longer, and the difference was statistically significant for *swipe* gestures overall (Table 6 in Appendix B.3). However, comparing the number of regions, we only found mid-air *swipe up* and *swipe front* gestures to be significantly longer for both 3 and 5 around-ear regions (Table 6). Mid-air *swipe up* gesture was also significantly longer for 7 around-ear regions. However, the gesture path length increase was not statistically significant for 3, 5, and 7 mid-air regions for *swipe back* and *swipe down*. A noticeable deviation for DV2 was observed for pinch gestures. Mid-air pinch gesture paths were significantly shorter than on-skin (Figure 4). Analyzing the number of regions, we observed significantly shorter mid-air path lengths for all numbers (3/5/7) of around-ear regions for *pinch in* gestures. Although the *pinch out* gesture paths were shorter for different numbers of mid-air around-ear regions, the differences were significant only for 7 region segmentations.

On-skin gesture accuracy (DV3) was significantly higher (Figure 4) than mid-air for *swipe up*, *swipe down*, *swipe front*, *pinch in*, *pinch out*, and *tap* gestures. Analyzing further individual gesture shapes and the number of around-ear regions, we observed that gesture accuracy for mid-air *swipe back* was better for all numbers of around-ear regions. The differences were statistically significant for both 3 and 7 region segmentation. For all other gestures and the number of regions, mid-air gestures were less accurate than on-skin. Overall (Figure 5 - all gestures), gesture accuracy remained relatively similar for 3 around-ear regions and starts diverging as the number of regions increases in mid-air and on-skin spaces. Figure 5 highlights the statistically significant differences in DV3 metric, and Table 8 in Appendix B.4 outlines associated statistical test results.

To summarize, The effect of interaction space choice is less prominent for fewer (< 5) around-ear regions - the performance starts to significantly degrade for mid-air gestures compared to on-skin gestures as the number of gesture regions increases (≥ 5). For a lower number of regions ($= 3$), mid-air and on-skin gestures have overall similar performances, with mid-air gestures being faster than (and as accurate as) on-skin gestures while trading off subtlety in terms of larger gesture motion (higher DV2).

5.2 Effect of Increasing Gesture Regions on Performance Metrics (RQ2)

We found that increasing both on-skin and mid-air gesture regions for gesture reuse had a significant effect on both *Gesture Time* (DV1) (Table 5) and *Gesture Accuracy* (DV3) (Table 9) across all gesture classes (Figure 5 - “All Gestures”). However, we did not find significant changes in *Gesture Path Length* (DV2) for overall gestures in the on-skin interaction space. We did observe a significant drop in path length (Table 7) when mid-air, around-ear regions increased from 3 to 5. However, the changes from 5 to 7 regions, or 3 to 7 regions, were insignificant.

Analyzing RQ2 for gesture time (DV1) by individual on-skin gesture class, we observed that gesture time increase was insignificant for on-skin *swipe down* and *pinch in* gestures when gesture regions increased from 3 to 5. However, the increase in gesture time was significant for all other gesture region increases (3 to 5, 5 to 7, or 3 to 7) for all other gesture classes. In comparison, gesture time significantly increases for all mid-air gesture classes when mid-air gesture regions increase from 3 to 5 and 3 to 7. Except for mid-air *swipe up*, *pinch in*, and *pinch out* gestures, increasing the number of gesture regions from 5 to 7 was also significant for all other gestures. This indicated a potential upper bound in terms of *gesture time* (DV2) for mid-air (3 regions) and on-skin (5-region) spaces for gesture reuse. For gesture path length (DV2), we found a significant effect only for mid-air *pinch out* gestures when considering the effect of increasing around-ear regions on individual gesture classes (Table 7). There was no significant effect on DV2 for all other mid-air and on-skin gestures. For *pinch out*, DV2 significantly increased from 5 to 7 segmentation.

Analyzing gesture accuracy (DV3) for individual mid-air and on-skin gesture classes, we observed a significant decrease in accuracy for all mid-air and on-skin gesture classes when the number of around-ear regions was increased from 3 to 7. Although gesture accuracy decreased for all gesture classes on increasing on-skin regions from 3 to 5, only *swipe back* and *pinch in* gestures showed a significant decrease. Gesture accuracy further decreased when the number of on-skin regions was increased from 5 to 7, and except for *swipe back*, all decreases were significant. When mid-air gesture regions increased from 3 to 5, gesture accuracy dropped for all gesture classes, and except for *pinch in*, all accuracy drops were significant. However, upon increasing mid-air regions further to 7, the drop in gesture accuracy was insignificant for most gesture classes except *swipe down* and *tap*. Table 9 in Appendix B.4 shows the statistical test summary for testing individual gesture classes for DV3 metric.

The above metrics indicate that gesture performance does not vary significantly when the number of around-ear, on-skin regions increase from 3 to 5. However, further on-skin region segmentation for gesture reuse can significantly degrade on-skin gesture performance. In contrast, overall mid-air gesture performance degrades significantly when mid-air regions increase from 3 to 5. However, further segmentation of mid-air gesture space does not significantly affect overall gesture performance. Overall, this indicates a threshold of 3 mid-air regions and 5 on-skin regions for supporting gesture reuse.

6 QUALITATIVE OBSERVATIONS

6.1 Best and Worst Regions

To answer RQ3 outlined in Section 3.1, we analyzed the gesture region boundaries collected in Phase 2 of our study (Section 4.4.2). Figure 6 outlines the most (*green*) and least (*red*) preferred regions across 3, 5, and 7 gesture regions in on-skin and mid-air interaction spaces, categorized by individual gesture classes (*swipe/pinch/tap*). The darker shades indicate where participants’ preferences for the most and least preferred regions overlap⁵.

Participants preferred to limit the gesture region within the frontal or peripheral view as much as possible for both on-skin and mid-air interaction spaces. Overlap of best and worst regions, alongside participant discussions, revealed some noticeable differences between mid-air and on-skin gesture region positioning. For on-skin gestures,

⁵ Figures represent a participant creating a uni-manual gesture with the left (non-dominant) hand. Side views represent left-hand side views for participants. Participants use the left hand to create uni-manual gestures on the side closest to hand used to create gestures. Section 4.2 provides details on non-dominant hand usage.

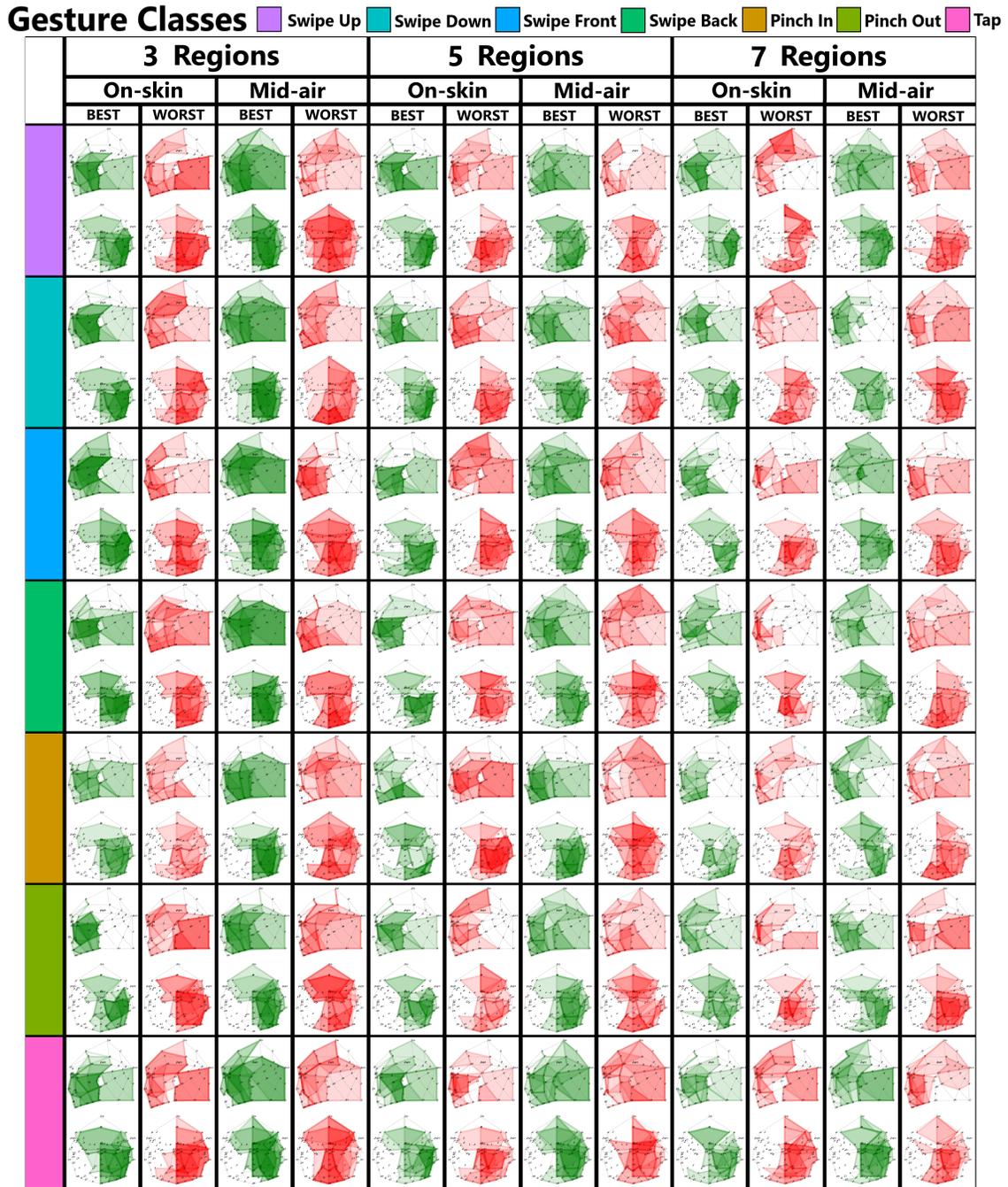


Fig. 6. Most & least preferred off-device, around-ear on-skin and mid-air gesture regions⁵ selected by all 18 participants. Green = Most Preferred and Red = Least Preferred. Darker areas indicate higher overlaps.

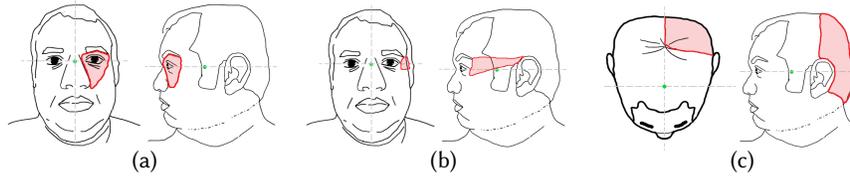


Fig. 7. Highlighted areas⁵ around different parts of the face for avoiding on-skin gestures.

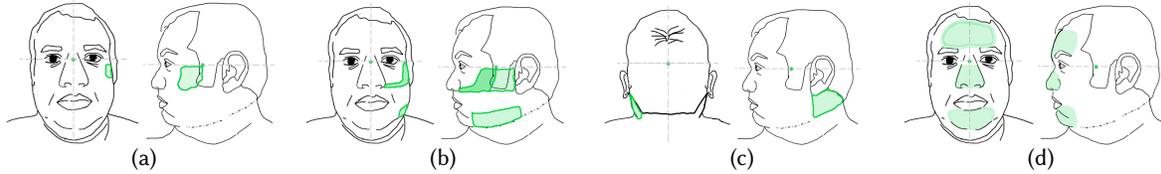


Fig. 8. Highlighted preferred areas⁵ around different parts of the face for on-skin gestures.

user-defined gesture region boundaries usually follow along the contour lines of the face. Participants showed considerable dislike for interacting with regions around the eyes. The area bounded by the *eyebrow*, *inner canthi* (corner of the eye on the side of the nose), *nasal ala*, and *outer canthi* (other corner of the eye near the temple) shown in Figure 7a are either not considered within different gesture regions or are usually considered among the worst-ranked regions for different gestures by almost all of the participants. Participants tended to avoid this area to protect the eye from accidental finger poking, keeping the frontal vision unblocked, and reducing Potential manual interaction conflict with eyewear. Another region of interest for on-skin gestures spans around the *temple* from the *outer canthi* to the top of the outer ear (Figure 7b). Analysis of on-skin region boundaries shows that some participants avoided this area or rated it less desirable than other on-skin regions. These participants were regular or occasional eyewear users; and were concerned about potential frame displacement from manual interaction with the temple area. However, some non-eyewear user participants identified this region as suitable for on-skin *tap* and horizontal *swipe* gestures. The area above and behind the *outer ear* (Figure 7c) is generally avoided for all on-skin interactions across 3/5/7 region segmentation. Participants report increased hand and arm joint bending compared to gestures within the frontal view, peripheral view, and under the *outer ear*. However, participants preferred these areas in Figure 7c for on-skin interaction over the area around the *eye* and *temple* (Figures 7a and 7b) as the inconvenience of resistance presented by hair region is less than an obstacle to vision or eyewear.

The highlighted area around the *cheekbone* (Figure 8a) was overwhelmingly preferred for tap and vertical swipe starting and ending points. Participants preferred the rigidity of the *cheekbone* for tactile feedback, and the gestures mentioned above in this highlighted region were highly accurate, regardless of the number of around-ear regions. Also, compared to other gesture regions with facial hair, this area offers less resistance to on-skin interaction, making it a viable candidate for mapping frequent tasks supported by the aforementioned on-skin gestures. The highlighted areas around the *cheekbone* and *jawline* in Figure 8b are preferred across all number of regions for on-skin horizontal pinch and swipe gestures, with the area around the *cheekbone* having higher preference than the *jawline*.

Participants do not generally prefer the area under the ear (Figure 8c) for reusing gestures in 3 on-skin regions. Nevertheless, for 5 and 7 regions, this area provides the participants with another convenient spot for vertical swipe, pinch, and tap gestures. Participants reported uncomfortable hand and arm joint bending in this area for horizontal on-skin swipe and pinch gestures. This indicates that on-skin interaction with this area should be limited to vertical motion gestures and taps. Like the area under the ear, highlighted areas in Figure 8d within

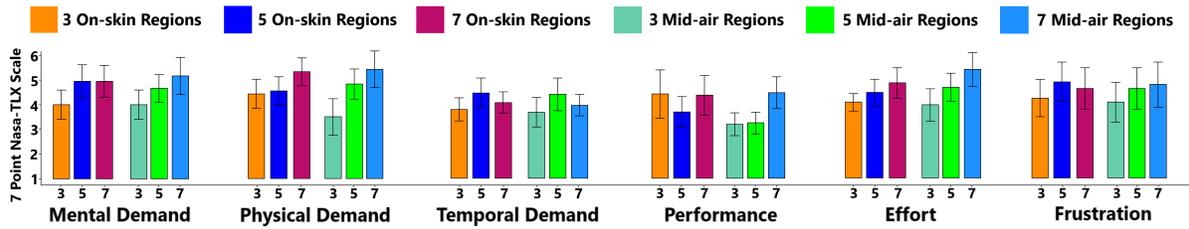


Fig. 9. NASA TLX survey results for comparing increasing gesture region segments across mid-air and on-skin around-ear interaction space, ignoring gesture types (with 95% confidence interval error bars)

the frontal view regions provided convenient, segmented regions for on-skin interaction. All highlighted areas (i.e., *chin*, *nose*, *forehead*) provided sufficient spaces for supporting on-skin swipes and taps. Participants reported horizontal pinches as more natural, subtle gestures suited for public spaces than vertical pinches for the above areas. Facial landmarks and contours in these highlighted regions make it easier to define region boundaries and have better gesture accuracy.

For mid-air gestures, gesture regions are larger with minimal obstacles than on-skin space. However, the lack of tactile feedback results in relying solely on frontal and peripheral views for feedback. For gestures outside of the field of view, participants solely rely on muscle/joint-based perception of the location of the hand and finger relative to the outer ear. Overall, participants are comfortable making uni-manual gesture motions in most of the mid-air space around the head except for the top and back. Due to uncomfortable hand and arm joint bending and site positioning outside the field of view, most participants do not include these two areas within the consideration of gesture region boundary definition or mark these as the least preferred areas. Another reported area of interest in mid-air gesture space is in front of the eyes within a direct field of view. Most participants felt more comfortable performing gestures outside this region but within peripheral view (over areas like the *cheek*, *jawline*, or *temple*) to avoid blocking the vision. Unlike on-skin gestures, participants were not wholly averted to make mid-air gestures directly above the area around the eyes highlighted in Figure 7a within frontal view. Participants reported eyewear not being a significant roadblock to performing mid-air uni-manual gestures, as gestures tended not to displace the eyewear frames. Although outside of the peripheral view, the area around the ear (directly over/ above/ below/ behind/ in-front) was preferred. Participants reported a better spatial sense of the relative location of the hand to the *ear*, especially when wearing an earable device such as a wireless earbud. They also explained that such uni-manual motions around the ear mimic interacting with an earable device. They are considered subtle and natural motion in public spaces due to the popularity of earbud usage in public and private spaces.

6.2 NASA-TLX Survey Results

During the study, participants completed a 7-point NASA TLX survey (Figure 9) for each specific interaction space and region segmentation to measure their perceived workload. This provided insight into changes in the perceived workload for increasing around-ear regions for gesture repetition.

Mental demand increased steadily when the number of around-ear regions increased from 3 to 5 for mid-air and on-skin gestures, plateauing for further on-skin region increase (to 7). In contrast, it steadily increased from 5 to 7 mid-air gesture region increase. Similar to *mental demand*, *physical demand* for mid-air gestures increased when the number of mid-air gestures was increased from 3 to 5 and from 5 to 7. Reported *physical demand* for reusing on-skin gestures across 3 and 5 segments was relatively similar. However, it considerably increased when the number of regions increased from 5 to 7. On-skin and mid-air gestures reported similar *temporal demand*

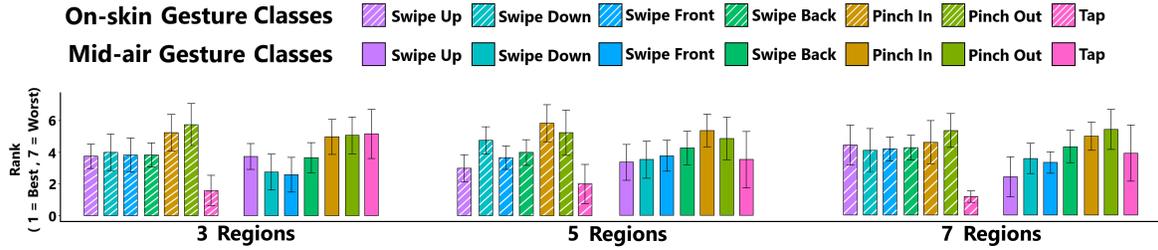


Fig. 10. Gesture rank for each block (with 95% confidence interval error bars).

patterns. Increased time pressure for reusing gestures across 5 gesture regions for both interaction spaces by the participants compared to 3/7 gesture regions.

Participants reported that their perceived *performance* was similar for 3 and 5 mid-air regions around the ear and considerably better than their on-skin counterparts. Surprisingly, for on-skin space, self-reported success in performing gestures over the intended region was better for 5 regions than 3 regions. The perceived *performance* was considerably worse for 7 regions and relatively similar for mid-air and on-skin. The participants reported a steady increase in both on-skin and mid-air gesture *effort* when the number of around-ear gesture regions increased. Similarly, *frustration* increased steadily with the number of mid-air around-ear gesture regions. However, *frustration* increased from 3 to 5 on-skin regions and slightly reduced when on-skin regions increased to 7, remaining higher than frustration for gestures in 3 on-skin regions. Moreover, the self-reported effort showed a steady increase for on-skin interaction space, but interestingly, frustration slightly reduced from 5 to 7 segments. Participants attributed this reduction in on-skin frustration to tactile feedback in finding target regions for more regions. Overall, these workload results support the threshold findings for mid-air and on-skin segmentations in Section 5.2.

6.3 Reported Gesture Ranking

Participants ranked each of the 7 gestures against each other for different numbers of mid-air, and on-skin around-ear regions. Figure 10 shows each block's self-reported mean ranking of different gestures. In general, participants reported that touch sensation helped identify the position of on-skin gestures for the target gesture regions, even when the regions were outside of the direct or peripheral view of the participants.

We found that *tap* remained the most preferred gesture for on-skin interaction space as the number of around-ear gesture regions increased. For mid-air gestures, *swipe down* and *swipe front* gestures received the best ranking for 3 regions. For 5 mid-air regions, although *tap* received the best ranking, it was closely followed by *swipe up*, *swipe down*, and *swipe front* gestures. When the number of around-ear mid-air gesture regions increased to 7, the ranking for *tap* fell behind *swipe up*, *swipe down*, and *swipe front* gestures.

Analysis of *Vicon* motion data and observation of participants' gesture motion show that all on-skin and mid-air *tap* gestures started 35 - 40 cm away from the head in the mid-air position within the visual space. As *tap* gestures approached the head, they took up smaller visual space in central or peripheral vision than other gesture classes. Skin touch sensory feedback at the end of on-skin *tap* gestures allowed the participant to quickly recognize if the gesture was over the correct gesture region. Also, motion data analysis showed that *tap* gestures involved fewer steps than *swipe* or *pinch* gestures. The former involved pulling the index finger towards a target region on/over the face and rapidly pulling away the hand. Compared to this motion, *swipe*/*pinch* gestures involved moving towards a targeted region on/around the face, followed by moving the index finger(s) in a horizontal/vertical direction and then rapidly pulling away from the face. The skin touch feedback, alongside reduced movement

complexity, contributed to the *tap* gesture's ranking as the primary choice for all numbers of around-ear gesture regions on the skin.

The lack of sensory feedback from skin touch in mid-air *tap* gestures makes the participant solely reliant on frontal or peripheral vision to perceive *tap* accuracy. This, coupled with the more prominent horizontal/vertical *swipe* motions over large mid-air regions, contributes to a better ranking of *swipe* over *tap* in 3 mid-air regions. With the increase in mid-air gesture regions, the maximum available area for each region starts to shrink. This results in smaller mid-air *swipe* motions being limited within the gesture boundary, providing it less visual prominence than similar *swipe* gestures for a smaller number of mid-air regions. The decrease in *swipe* prominence contributes to a slightly worsening ranking of *swipe* gestures in 5 mid-air regions compared to 3 regions, making all *swipe* and *tap* gestures receive similar rankings from participants. Vicon motion analysis also shows that for mid-air and on-skin *tap* gestures, gesture motion is usually bound within the same gesture region and does not mostly cross region boundaries from start to finish, making it difficult for participants to recognize mid-air *tap* gestures over incorrect regions until the motion is complete. In 7 mid-air regions, the reduction in maximum available area for each region, as well as the abovementioned *tap* motion pattern, contributes to increased gesture attempts (compared to 3 and 5 mid-air regions and other *swipe* gestures) over different mid-air regions to make corrections between the starting and the ending of the gesture recording window -resulting in the shift in self-reported rankings in figure 10, as well as increasing average path length for mid-air *tap* gestures in figure 4).

For mid-air horizontal/vertical *swipe* gestures, Vicon motion data analysis revealed larger motions compared to their on-skin counterparts (Figures 4 and 5). Mid-air *swipe up*, *swipe down*, and *swipe front* received comparable self-reported ranking than *tap* in the same interaction space (considerably better in 3 mid-air regions, and slightly better in 5 and 7 regions) as they started and ended within the direct or peripheral vision, usually crossing from one segmented region to another to reach the destination region. As such, participants were more confident that the *swipe* was done across the correct gesture region.

Participants reported discomfort in the backward bending of wrist and elbow joints for both on-skin and mid-air *swipe back* gestures. For example, "I feel very uncomfortable swiping my hand back compared to the other gestures." -P09. The participants did not report similar discomfort for other gesture types across both interaction spaces. Also, for mid-air *swipe back* gestures, the starting position was *mainly* within the direct/peripheral view of the participants. However, for gesture regions that went beyond the peripheral view of the participants, the ending of such gestures could not be visually inspected by the participants. For such regions, the lack of tactile feedback reduced confidence in the perceived performance of mid-air *swipe back* over the intended gesture region, leading to worse ranking compared to other mid-air gestures across all numbers of around-ear regions. However, this effect was not observed across different numbers of on-skin regions due to tactile feedback, and the gesture ranking of on-skin *swipes* was relatively similar.

Because of increased complexity, *pinch* gestures took considerably more time than other gestures across all numbers of around-ear regions. The relative complexity of gesture motions compared to *swipe* and *tap* gestures resulted in users preferring later gestures to *pinch* gestures in mid-air and on-skin regions. These reported rankings, alongside the gesture performance metrics discussed in Section 5, allow potential gesture designers to compare different gestures to choose them to map with appropriate tasks for different usage scenarios.

7 DISCUSSION

7.1 Interaction Space and Number of Gesture Regions

Our analysis shows that mid-air gestures were faster than on-skin gestures across 3, 5, and 7 around-ear regions (Figure 5). However, on-skin gestures demonstrated significantly better accuracy than mid-air gestures for 5 and 7 around-ear regions. For 3 regions around the ear, there was no discernible difference in performance between these two interaction spaces. When considering accuracy, response time, and self-perceived performance from

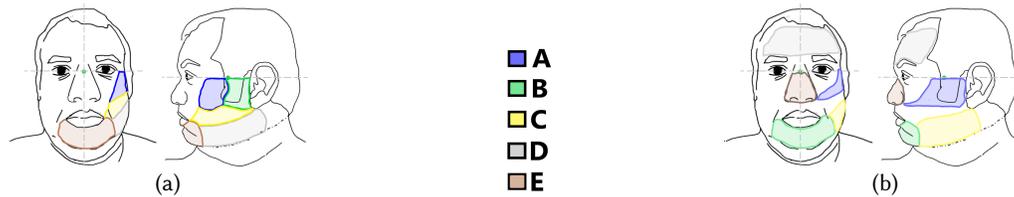


Fig. 11. Sample 5 on-skin gesture region boundary from two participants showing gesture region packing density.

the NASA-TLX rating, it becomes evident that the most optimal condition for conducting interactions is 3 mid-air regions. Several potential factors contribute to this efficiency gap.

- Mid-air gestures do not require physical contact with a surface, allowing for quicker initiation and execution.
- The absence of physical constraints associated with on-skin gestures provides more fluid and intuitive movement, enabling users to perform actions swiftly.
- The presence of larger region boundaries and more region location possibilities within the peripheral view compared to the same number of on-skin regions.

The following comments collected during the study revealed more insights into the preference for mid-air gestures compared to on-skin gestures for 3 gesture regions:

“For 3 Segmentation, I can create larger mid-air gesture regions than on-skin and avoid skin scraping.” -P08

However, as the number of around-ear gesture regions increased from 3 regions, participants started reporting a preference shift towards on-skin interaction space for most gestures because of tactile sensation from skin touches. Perceived workload metrics except self-reported performance become similar for 5 regions (Figure 9) for both interaction spaces. Although gestures over 5 mid-air regions are faster and have a higher self-reported preference, statistical analysis shows accuracy to be significantly better for 5 on-skin regions than mid-air counterparts (Figure 5). Figure 9 also shows a noticeable improvement in self-reported performance when the number of on-skin regions increases from 3 to 5, supporting the preference shift observation towards on-skin for 5 regions. When the number of gesture regions increases to 7, we observe a more significant drop in gesture accuracy and an increase in time for mid-air regions compared to on-skin regions. Overall results and observations from participant comments indicate that on-skin gestures are preferred over mid-air for 7 regions. The following comment by participants explains tactile feedback as a possible reason for the preference shift in interaction space for a higher number of gesture regions: *“It is difficult to gesture over the correct place when the number of regions increases to 5 and beyond. Touching the skin helps me to quickly identify if I am making a gesture over the correct region. In mid-air gestures, it is challenging to make gestures over the correct boundary with only visual feedback; even more so when gestures are in the air behind the ears.” -P10*

Overall, participants favored a 3-region configuration in both mid-air and on-skin interaction spaces and indicated that the configuration could be increased up to 5 regions for on-skin spaces because of tactile feedback. Participants did not prefer 7 regions in both interaction spaces. These findings are similar to the B2B [50] study analyzing gesture reuse over multiple regions for smartwatches, indicating that fewer regions can enhance interaction performance in terms of accuracy and time efficiency.

7.2 Observations on Gesture Region Size, Location and Boundaries

We observed participants using facial contour lines or buffer spaces for defining on-skin gesture space boundaries and tended to create loosely packed gesture regions whenever possible. In minor instances where participants did not create region boundaries along facial contour lines (e.g., dividing the jawline spanning from the end of the chin to the jaw angle into two regions), we observed incorrect swipe and pinch gestures as they could not

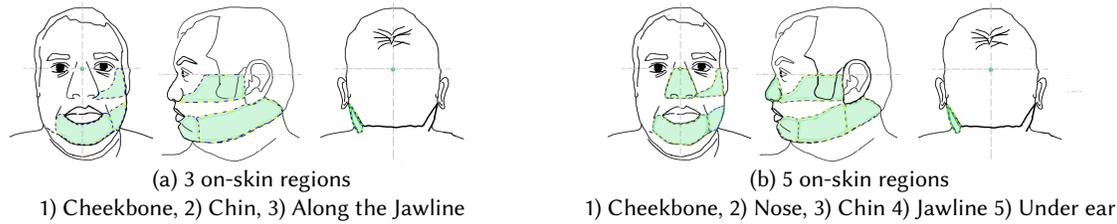


Fig. 12. Recommended areas⁵ for re-using gestures in 3 and 5 on-skin regions. Dotted lines indicate region boundaries.

correctly start or stop over the destined region. As on-skin gesture regions increased to 5 and 7, participants still found techniques to create loosely packed gesture regions and large gesture region boundaries with facial contours or buffer space separations by choosing areas outside the field of view. However, for on-skin gestures, gesture accuracy was more affected by gesture region packing structure and facial contour lines defining the gesture region boundaries than the size of the gesture region. We observed more incorrect gestures for densely packed gesture regions (e.g., multiple regions on the cheek) and region boundaries that did not follow facial contour lines. Figure 11a shows a densely packed on-skin gesture region structure. The regions in this structure were reused for *swipe back* and *swipe down* gestures by the participant. Although the participant could perform most *swipe back* gestures over all regions correctly, the *swipe down* gesture over region C incorrectly ended over region D over multiple attempts due to the region packing structure and shape of region C. The participant later remarked that during *swipe down* over the cheekbone (region A), gestures did not cross over to region C on the cheek as the region boundary (end of *Zygomatic bone*) was evident from tactile feedback, which was not the case for region C. Figure 11b shows another region packing structure created by another participant where buffer spaces exist between most gesture regions. Like Figure 11a, regions on the cheek and jawbone border each other, and the region boundary (*end of chin*) is prominent. These loose region packing structures reported higher overall accuracy than those in Figure 11a. We also observed a similar relationship between gesture performance and region packing structure relationship for 7 regions, where more regions were packed without facial contour lines as gesture region boundaries, resulting in increased incorrect gestures. These observations match the results obtained by Dezfuli et al. [12] for imaginary user interfaces, where on-skin gesture effectiveness decreases with increasing gesture region packing density and expand their gesture reuse findings from palm-based interfaces to earable devices.

We observed a similar compounding effect of region size, location, and facial landmarks on mid-air gesture performances. Smaller gesture regions within the field of view were overall more accurate than larger gesture regions outside the field of view (e.g., behind /above the ear). Within the field of view, facial areas such as the *nose*, *corner of the eye*, *corner of the mouth*, and *starting of the cheekbone* served as visuospatial referent points, and mid-air gestures closer to these points were more accurately performed over the correct region. As on-skin gestures started from mid-air, hand proprioception for the above-mentioned visuospatial referent points also positively impacted gesture performance for on-skin regions around these landmarks. These findings are similar to the finger-pointing study using visuospatial reference points by Gustafson et al. [20] for imaginary mid-air interfaces.

7.3 Implications for Design

Based on our preliminary insights into around-ear design space, we present the following design implications for implementing on-skin and mid-air, off-device gestures for earable devices:

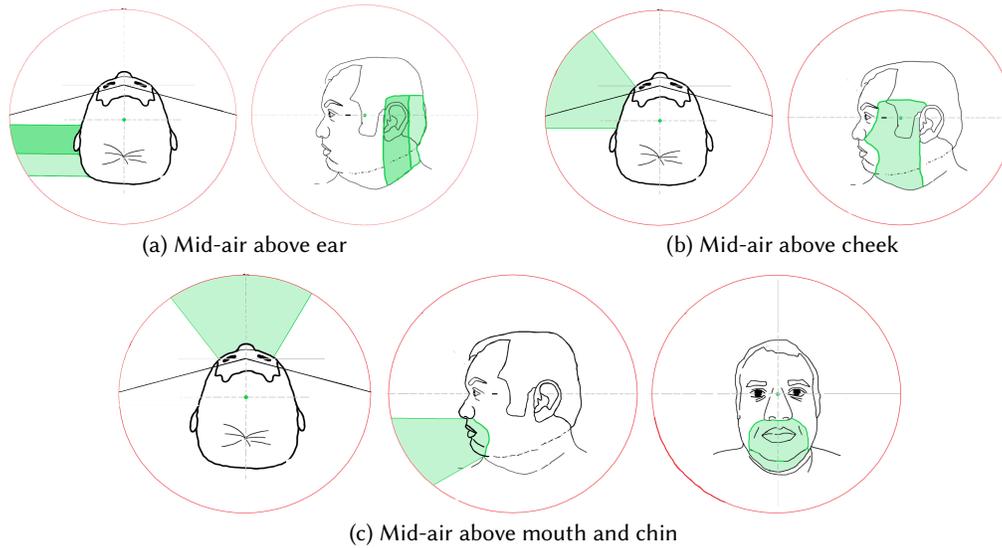


Fig. 13. Recommended areas⁵ for re-using gestures in 3 mid-air regions. The red circle represents DV1 measurement boundary (30cm) from the center of the head discussed in Section 4.2.

- The regions for on-skin and mid-air gesture reuse should be limited to 5 and 3, respectively, and mid-air gestures are preferred for 3 regions. Figures 12 (on-skin) and 13 (mid-air) propose approximate gesture regions based on Figure 6 and qualitative observations on region boundary patterns.
- For on-skin gestures, most frequent tasks should be mapped to the *cheekbone* area due to its location within peripheral vision within the field of view, reduced obstacle and friction from lack of facial hair, and being outside of eyewear regions and underlying bone support provides better tactile feedback.
- For mid-air gestures, the most frequent tasks should be mapped to the area from the *temple* to the *jaw* through the *cheek* area (Figure 13b), as uni-manual gesture motion with this region mimics natural hand-to-face interactions, and the location is within the peripheral view.
- The *temple* region (Figure 7b) should be avoided for on-skin gestures to avoid potential eyewear frame displacement issues. The highlighted region can be a buffer between on-skin gesture regions above and below the temple. However, such restriction does not apply to mid-air gestures.
- Designing touch gestures below the *eye* to the *corner of the mouth* (Figure 6a) should be avoided because of potential eyewear issues.
- Gesture regions should be loosely packed as much as possible, and facial landmarks and contour lines should be leveraged while defining gesture regions.

8 LIMITATIONS AND FUTURE WORK

We used motion tracking (Vicon camera array) to study around-ear gestures. However, exploring embedded earbud motion sensors for gesture recognition around the device could offer a more realistic approach. Our primary research focused on the exploration of reusing around-ear gestures by gesture space segmentation rather than the supporting technology itself. Notably, no commercially available Earable devices currently recognize both on-skin and mid-air unimanual gestures around the device simultaneously. Developing such a device for detecting gestures in these interaction spaces represents a compelling avenue for future research. In addition, We excluded interaction techniques on the ear itself, which was well-explored in previous research [36]. Adding

on-ear interaction could complicate our study with numerous combinations. Nevertheless, comparing around-ear and on-ear interactions could broaden insights into gesture-based Earable device interactions.

Due to the length of our study, we limited our gesture reuse exploration to only a sitting stance with the manual encumbrance of the dominant hand. However, it is necessary to understand participant preference for gesture space, number of regions, and region boundary definition in other real-world scenarios (e.g., standing, walking with the encumbered dominant and non-dominant hand, walking without manual encumbrance) for a more thorough understanding of around-ear unimanual gesture space for supporting off-device earable gestures in the real world. In the future, we plan to explore around-ear, off-device earable gesture reuse with the aforementioned gesture classes across 3 to 5 gesture regions to build upon our initial work and present a more comprehensive set of insights for earable interaction designers. We also plan to look into gesture-task mapping for proposed gesture reuse in various usage scenarios across *smart home environments*, *public display interactions*, and *AR/VR interactions*, among others. For example, an *on-cheek* region can control ceiling lights in a *smart home* environment. In contrast, a region on the *chin* can manage the TV, and the region around the *jawline* can interact with the smartphone. This can provide useful insight into the association of gesture region location with task importance in various scenarios and build upon our initial work for supporting proposed off-device earable interaction techniques in the real world.

The observations in Section 7.2 set apart proposed around-ear earable interaction techniques with other gesture-based interfaces operating entirely within the field of view. Fitt's law [18] assumes target visibility relates gesture region size directly with gesture performance. As evidenced by the example scenarios (Section 7.2), some target regions in our study fall outside the field of view, and gesture region size alone can not account for gesture performance for proposed around-ear interaction in earable devices. More recent complex models accounting for factors like location [7, 52] have been proposed to account for target regions outside the field of view. Ens [15] explores these extensions of Fitt's Law in a head-worn display (HWD) and the head-motion context in a user study. However, their findings can not be directly translated into an earable context where the head motion to target mid-air and on-skin spaces outside of field-of-view using unimanual gestures is impossible. More complex gesture performance models extending Fitt's Law addressing region size, location, tactile feedback (on-skin gesture), or hand proprioception (mid-air gestures) need to be developed in around-ear, earable contexts. As a first attempt to explore uni-manual around-ear off-device gestures for earables, this work focuses on investigating users' performances in different interaction spaces and how they generally vary for different numbers of regions. More in-depth study regarding the specific properties of the gesture region targets is left for future work.

9 CONCLUSION

In this paper, we experimented with gesture reuse for mid-air and on-skin-around-ear interaction space for Earable to increase the gestural input vocabulary. Such an approach could counter the limited on-device interaction space and avoid the pitfalls with other potential interaction spaces such as ear-bending, facial expression, head movements, or voice-based interactions. Our exploration found that tactile feedback switches user preference from mid-air to on-skin spaces for more gesture regions. For both spaces, we summarize that the interaction space should be segmented into at most 5 regions for gesture reuse, and for mid-air spaces, 3 regions are preferred. Besides our observations concerning gesture class and location preferences across different numbers of around-ear regions, our gesture motion analysis study revealed the overlapping area of preference for an increasing number of regions for different gestures, along with a heatmap of the most and least preferred regions around the ear - providing the ubicomp researcher and interaction designer communities a roadmap to map the most frequent tasks performed with Earable to certain gesture regions in mid-air and on-skin space. Although our work is by no means an exhaustive study of gesture reuse for off-device earable interaction, further research in this direction

could complement existing research into earable gesture recognition technology to make the proposed off-device interaction market ready for commercialization and popular among end users.

REFERENCES

- [1] David Ahlström, Khalad Hasan, and Pourang Irani. 2014. Are You Comfortable Doing That? Acceptance Studies of around-Device Gestures in and for Public Settings. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (Toronto, ON, Canada) (*MobileHCI '14*). Association for Computing Machinery, New York, NY, USA, 193–202. <https://doi.org/10.1145/2628363.2628381>
- [2] Khaled Alkiek, Khaled A. Harras, and Moustafa Youssef. 2022. EarGest: Hand Gesture Recognition with Earables. In *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE Press, Stockholm, Sweden, 91–99. <https://doi.org/10.1109/SECON55815.2022.9918622>
- [3] Takashi Amesaka, Hiroki Watanabe, and Masanori Sugimoto. 2019. Facial Expression Recognition Using Ear Canal Transfer Function. In *Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) (*ISWC '19*). Association for Computing Machinery, New York, NY, USA, 1–9. <https://doi.org/10.1145/3341163.3347747>
- [4] Apple. 2023. AirPods Pro (2nd generation) - technical specifications. <https://www.apple.com/ca/airpods-pro/specs/>.
- [5] Felix Barroso, Norbert Freedman, and Stanley Grand. 1980. Self-touching, performance, and attentional processes. *Perceptual and Motor Skills* 50, 3, suppl (1980), 1083–1089. <https://doi.org/10.2466/pms.1980.50.3c.1083>
- [6] Ana Berdasco, Gustavo López, Ignacio Diaz, Luis Quesada, and Luis A Guerrero. 2019. User experience comparison of intelligent personal assistants: Alexa, Google Assistant, Siri and Cortana. *Multidisciplinary Digital Publishing Institute Proceedings* 31, 1 (2019), 51. <https://doi.org/10.3390/proceedings2019031051>
- [7] Xiang Cao, Jacky Jie Li, and Ravin Balakrishnan. 2008. Peephole Pointing: Modeling Acquisition of Dynamically Revealed Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) (*CHI '08*). Association for Computing Machinery, New York, NY, USA, 1699–1708. <https://doi.org/10.1145/1357054.1357320>
- [8] Yetong Cao, Chao Cai, Anbo Yu, Fan Li, and Jun Luo. 2023. EarAcE: Empowering Versatile Acoustic Sensing via Earable Active Noise Cancellation Platform. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 2, Article 47 (jun 2023), 23 pages. <https://doi.org/10.1145/3596242>
- [9] Yu-Chun Chen, Chia-Ying Liao, Shuo-wen Hsu, Da-Yuan Huang, and Bing-Yu Chen. 2020. Exploring User Defined Gestures for Ear-Based Interactions. *Proc. ACM Hum.-Comput. Interact.* 4, ISS, Article 186 (nov 2020), 20 pages. <https://doi.org/10.1145/3427314>
- [10] Seokmin Choi, Yang Gao, Yincheng Jin, Se jun Kim, Jiyang Li, Wenyao Xu, and Zhanpeng Jin. 2022. PPGface: Like What You Are Watching? Earphones Can "Feel" Your Facial Expressions. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 48 (jul 2022), 32 pages. <https://doi.org/10.1145/3534597>
- [11] Romit Roy Choudhury. 2021. Earable Computing: A New Area to Think About. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications* (Virtual, United Kingdom) (*HotMobile '21*). Association for Computing Machinery, New York, NY, USA, 147–153. <https://doi.org/10.1145/3446382.3450216>
- [12] Niloofar Dezfuli, Mohammadreza Khalilbeigi, Jochen Huber, Florian Müller, and Max Mühlhäuser. 2012. PalmRC: Imaginary Palm-Based Remote Control for Eyes-Free Television Interaction. In *Proceedings of the 10th European Conference on Interactive TV and Video* (Berlin, Germany) (*EuroITV '12*). Association for Computing Machinery, New York, NY, USA, 27–34. <https://doi.org/10.1145/2325616.2325623>
- [13] John D Eastwood, Daniel Smilek, and Philip M Merikle. 2003. Negative facial expression captures attention and disrupts performance. *Perception & psychophysics* 65, 3 (2003), 352–358. <https://doi.org/10.1080/026999300378996>
- [14] Paul Ekman and Wallace V. Friesen. 2006. Hand Movements. *Journal of Communication* 22, 4 (02 2006), 353–374. <https://doi.org/10.1111/j.1460-2466.1972.tb00163.x>
- [15] Barrett Ens, David Ahlström, and Pourang Irani. 2016. Moving Ahead with Peephole Pointing: Modelling Object Selection with Head-Worn Display Field of View Limitations. In *Proceedings of the 2016 Symposium on Spatial User Interaction* (Tokyo, Japan) (*SUI '16*). Association for Computing Machinery, New York, NY, USA, 107–110. <https://doi.org/10.1145/2983310.2985756>
- [16] Jonny Farrington, Vanessa Oni, Chi Ming Kan, and Leo Poll. 1999. Co-Modal Browser-An Interface for Wearable Computers. In *Proceedings of the 3rd IEEE International Symposium on Wearable Computers* (*ISWC '99*). IEEE Computer Society, USA, 45. <https://doi.org/10.1109/ISWC.1999.806644>
- [17] Andrea Ferlini, Alessandro Montanari, Cecilia Mascolo, and Robert Harle. 2020. Head Motion Tracking Through In-Ear Wearables. In *Proceedings of the 1st International Workshop on Earable Computing* (London, United Kingdom) (*EarComp'19*). Association for Computing Machinery, New York, NY, USA, 8–13. <https://doi.org/10.1145/3345615.3361131>
- [18] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381. <https://doi.org/10.1037/0096-3445.121.3.262>
- [19] Hyunjae Gil and Ian Oakley. 2023. ThumbAir: In-Air Typing for Head Mounted Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 164 (jan 2023), 30 pages. <https://doi.org/10.1145/3569474>

- [20] Sean Gustafson, Daniel Bierwirth, and Patrick Baudisch. 2010. Imaginary Interfaces: Spatial Interaction with Empty Hands and without Visual Feedback. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) (*UIST '10*). Association for Computing Machinery, New York, NY, USA, 3–12. <https://doi.org/10.1145/1866029.1866033>
- [21] Jinni A Harrigan, John R Kues, John J Steffen, and Robert Rosenthal. 1987. Self-touching and impressions of others. *Personality and Social Psychology Bulletin* 13, 4 (1987), 497–512. <https://doi.org/10.1177/0146167287134007>
- [22] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-Air Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (*CHI '14*). Association for Computing Machinery, New York, NY, USA, 1063–1072. <https://doi.org/10.1145/2556288.2557130>
- [23] Tamzid Hossain, Md. Fahimul Islam, William Delamare, Farida Chowdhury, and Khalad Hasan. 2022. Exploring Social Acceptability and Users' Preferences of Head-and Eye-Based Interaction with Mobile Devices. In *Proceedings of the 20th International Conference on Mobile and Ubiquitous Multimedia* (Leuven, Belgium) (*MUM '21*). Association for Computing Machinery, New York, NY, USA, 12–23. <https://doi.org/10.1145/3490632.3490636>
- [24] Tahera Hossain, Md Shafiqul Islam, Md Atiqur Rahman Ahad, and Sozo Inoue. 2019. Human Activity Recognition Using Earable Device. 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, 81–84. <https://doi.org/10.1145/3341162.3343822>
- [25] Imagimob. 2021. AOI A1 Touchless In-Ear Headphones. <https://www.imagimob.com/gestures>
- [26] Delaram Jarchi, Charence Wong, Richard Mark Kwasnicki, Ben Heller, Garry A. Tew, and Guang-Zhong Yang. 2014. Gait Parameter Estimation From a Miniaturized Ear-Worn Sensor Using Singular Spectrum Analysis and Longest Common Subsequence. *IEEE Transactions on Biomedical Engineering* 61, 4 (2014), 1261–1273. <https://doi.org/10.1109/TBME.2014.2299772>
- [27] Yincheng Jin, Yang Gao, Xuhai Xu, Seokmin Choi, Jiyang Li, Feng Liu, Zhengxiong Li, and Zhanpeng Jin. 2022. EarCommand: "Hearing" Your Silent Speech Commands In Ear. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 57 (jul 2022), 28 pages. <https://doi.org/10.1145/3534613>
- [28] Yincheng Jin, Yang Gao, Yanjun Zhu, Wei Wang, Jiyang Li, Seokmin Choi, Zhangyu Li, Jagmohan Chauhan, Anind K. Dey, and Zhanpeng Jin. 2021. SonicASL: An Acoustic-Based Sign Language Gesture Recognizer Using Earphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 67 (jun 2021), 30 pages. <https://doi.org/10.1145/3463519>
- [29] Vimal Kakaraparthi, Qijia Shao, Charles J. Carver, Tien Pham, Nam Bui, Phuc Nguyen, Xia Zhou, and Tam Vu. 2021. FaceSense: Sensing Face Touch with an Ear-Worn System. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 110 (sep 2021), 27 pages. <https://doi.org/10.1145/3478129>
- [30] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Alessandro Montanari. 2018. Earables for Personal-Scale Behavior Analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89. <https://doi.org/10.1109/MPRV.2018.03367740>
- [31] Takashi Kikuchi, Yuta Sugiura, Katsutoshi Masai, Maki Sugimoto, and Bruce H. Thomas. 2017. EarTouch: turning the ear into an input surface. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) (*MobileHCI '17*). Association for Computing Machinery, New York, NY, USA, Article 27, 6 pages. <https://doi.org/10.1145/3098279.3098538>
- [32] Marion Koelle, Swamy Ananthanarayan, and Susanne Boll. 2020. Social Acceptability in HCI: A Survey of Methods, Measures, and Design Strategies. 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, 1–19. <https://doi.org/10.1145/3313831.3376162>
- [33] Maurice H Krout. 1954. An Experimental Attempt to Produce Unconscious Manual Symbolic Movements. *The Journal of General Psychology* 51, 1 (1954), 93–120. <https://doi.org/10.1080/00221309.1954.9920209>
- [34] Yen Lee Angela Kwok, Jan Gralton, and Mary-Louise McLaws. 2015. Face touching: A frequent habit that has implications for hand hygiene. *American Journal of Infection Control* 43, 2 (2015), 112–114. <https://doi.org/10.1016/j.ajic.2014.10.015>
- [35] 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>
- [36] Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, and Max Mühlhäuser. 2013. EarPut: Augmenting behind-the-Ear Devices for Ear-Based Interaction. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) (*CHI EA '13*). Association for Computing Machinery, New York, NY, USA, 1323–1328. <https://doi.org/10.1145/2468356.2468592>
- [37] Jindong Liu, Edward Johns, Louis Atallah, Claire Pettitt, Benny Lo, Gary Frost, and Guang-Zhong Yang. 2012. An Intelligent Food-Intake Monitoring System Using Wearable Sensors. In *Proceedings of the 2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks* (BSN '12). IEEE, USA, 154–160. <https://doi.org/10.1109/BSN.2012.11>
- [38] Katsutoshi Masai, Kai Kunze, Daisuke Sakamoto, Yuta Sugiura, and Maki Sugimoto. 2020. Face Commands - User-Defined Facial Gestures for Smart Glasses. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, USA, 374–386. <https://doi.org/10.1109/ISMAR50242.2020.00064>

- [39] Douglas McIlwraith, Julien Pansiot, and Guang-Zhong Yang. 2010. Wearable and ambient sensor fusion for the characterisation of human motion. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Taipei, Taiwan, 5505–5510. <https://doi.org/10.1109/IROS.2010.5650512>
- [40] Christian Metzger, Matt Anderson, and Thad Starner. 2004. FreeDigiter: A Contact-Free Device for Gesture Control. In *Proceedings of the Eighth International Symposium on Wearable Computers (ISWC '04)*. IEEE Computer Society, USA, 18–21. <https://doi.org/10.1109/ISWC.2004.23>
- [41] Chulhong Min, Akhil Mathur, and Fahim Kawsar. 2018. Exploring Audio and Kinetic Sensing on Earable Devices. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications (Munich, Germany) (WearSys '18)*. Association for Computing Machinery, New York, NY, USA, 5–10. <https://doi.org/10.1145/3211960.3211970>
- [42] Stephanie Margarete Mueller, Sven Martin, and Martin Grunwald. 2019. Self-touch: contact durations and point of touch of spontaneous facial self-touches differ depending on cognitive and emotional load. *PLoS one* 14, 3 (2019), e0213677. <https://doi.org/10.1371/journal.pone.0213677>
- [43] Miguel A. Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. 2013. Memorability of Pre-Designed and User-Defined Gesture Sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Paris, France) (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 1099–1108. <https://doi.org/10.1145/2470654.2466142>
- [44] Colver Ken Howe Ne, Jameel Muzaffar, Aakash Amlani, and Manohar Bance. 2021. Hearables, in-ear sensing devices for bio-signal acquisition: a narrative review. *Expert Review of Medical Devices* 18, sup1 (2021), 95–128. <https://doi.org/10.1080/17434440.2021.2014321>
- [45] Mark Nicas and Daniel Best. 2008. A study quantifying the hand-to-face contact rate and its potential application to predicting respiratory tract infection. *Journal of occupational and environmental hygiene* 5, 6 (2008), 347–352. <https://doi.org/10.1080/15459620802003896>
- [46] Joseph Plazak and Marta Kersten-Oertel. 2018. A Survey on the Affordances of “Hearables”. *Inventions* 3, 3 (2018), 48. <https://doi.org/10.3390/inventions3030048>
- [47] Hanae Rateau, Edward Lank, and Zhe Liu. 2022. Leveraging Smartwatch and Earbuds Gesture Capture to Support Wearable Interaction. *Proc. ACM Hum.-Comput. Interact.* 6, ISS, Article 557 (nov 2022), 20 pages. <https://doi.org/10.1145/3567710>
- [48] Grand View Research. 2022. Smart Headphones Market Analysis By Product (Wired, Wireless) And Segment Forecasts To 2022. <https://www.grandviewresearch.com/industry-analysis/smart-headphones-market>.
- [49] Grand View Research. 2023. Earphones And Headphones Market Size, Share & Trends Analysis Report By Price Band (>100, 50-100, <50), By Product (Earphones And Headphones), By Technology, By Application, By Region, And Segment Forecasts, 2023 - 2030. <https://www.grandviewresearch.com/industry-analysis/earphone-and-headphone-market>.
- [50] Bradley Rey, Kening Zhu, Simon Tangi Perrault, Sandra Bardot, Ali Neshati, and Pourang Irani. 2022. Understanding and Adapting Bezel-to-Bezel Interactions for Circular Smartwatches in Mobile and Encumbered Scenarios. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 201 (sep 2022), 28 pages. <https://doi.org/10.1145/3546736>
- [51] Tobias Röddiger, Christopher Clarke, Paula Breitling, Tim Schneegans, Haibin Zhao, Hans Gellersen, and Michael Beigl. 2022. Sensing with Earables: A Systematic Literature Review and Taxonomy of Phenomena. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 135 (sep 2022), 57 pages. <https://doi.org/10.1145/3550314>
- [52] Michael Rohs and Antti Oulasvirta. 2008. Target Acquisition with Camera Phones When Used as Magic Lenses. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Florence, Italy) (CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1409–1418. <https://doi.org/10.1145/1357054.1357275>
- [53] Samsung. 2023. Samsung Galaxy Buds Pro. <https://www.samsung.com/global/galaxy/galaxy-buds-pro/specs/>.
- [54] Khairul Khaizi Mohd Shariff, Auni Nadiyah Yusni, Mohd Adli Md Ali, Megat Syahirul Amin Megat Ali, Megat Zuhairy Megat Tajuddin, and MAA Younis. 2022. CW Radar Based Silent Speech Interface Using CNN. In *2022 IEEE Symposium on Wireless Technology & Applications (ISWTA)*. IEEE, USA, 76–81. <https://doi.org/10.1109/ISWTA55313.2022.9942730>
- [55] Fangli Song and Wei Wang. 2021. Designing the Security Enhancement Features in the Future Headphone Experience. In *Design, User Experience, and Usability: Design for Contemporary Technological Environments: 10th International Conference, DUXU 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part III*. Springer, Springer-VerlagBerlin, Heidelberg, 90–108. https://doi.org/10.1007/978-3-030-78227-6_8
- [56] Xingzhe Song, Kai Huang, and Wei Gao. 2022. FaceListener: Recognizing Human Facial Expressions via Acoustic Sensing on Commodity Headphones. In *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, USA, 145–157. <https://doi.org/10.1109/IPSN54338.2022.00019>
- [57] Tanmay Srivastava, Prerna Khanna, Shijia Pan, Phuc Nguyen, and Shubham Jain. 2022. MuteIt: Jaw Motion Based Unvoiced Command Recognition Using Earable. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 140 (sep 2022), 26 pages. <https://doi.org/10.1145/3550281>
- [58] Dennis Stanke, Pia Brandt, and Michael Rohs. 2022. Exploring the Design Space of Headphones as Wearable Public Displays. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 295, 7 pages. <https://doi.org/10.1145/3491101.3519756>

- [59] Emi Tamaki, Takashi Miyaki, and Jun Rekimoto. 2009. Brainy Hand: An Ear-Worn Hand Gesture Interaction Device. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 4255–4260. <https://doi.org/10.1145/1520340.1520649>
- [60] Jean Vanderdonckt, Nathan Magrofuoco, Suzanne Kieffer, Jorge Pérez, Ysabelle Rase, Paolo Roselli, and Santiago Villarreal. 2019. Head and Shoulders Gestures: Exploring User-Defined Gestures with Upper Body. In *Design, User Experience, and Usability. User Experience in Advanced Technological Environments: 8th International Conference, DUXU 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part II* (Orlando, FL, USA). Springer-Verlag, Berlin, Heidelberg, 192–213. https://doi.org/10.1007/978-3-030-23541-3_15
- [61] Pui Chung Wong, Kening Zhu, Xing-Dong Yang, and Hongbo Fu. 2020. Exploring Eyes-Free Bezel-Initiated Swipe on Round Smartwatches. 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, 1–11. <https://doi.org/10.1145/3313831.3376393>
- [62] Xuhai Xu, Haitian Shi, Xin Yi, WenJia Liu, Yukang Yan, Yuanchun Shi, Alex Mariakakis, Jennifer Mankoff, and Anind K. Dey. 2020. EarBuddy: Enabling On-Face Interaction via Wireless Earbuds. 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, 1–14. <https://doi.org/10.1145/3313831.3376836>
- [63] Hoon Sik Yoo and Da Young Ju. 2019. Analysis of the emotional user experience elements of wireless earphones. In *Advances in Interdisciplinary Practice in Industrial Design: Proceedings of the AHFE 2018 International Conference on Interdisciplinary Practice in Industrial Design, July 21–25, 2018, Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida, USA 9*. Springer, USA, 247–255. https://doi.org/10.1007/978-3-319-94601-6_26
- [64] Shang Zeng, Haoran Wan, Shuyu Shi, and Wei Wang. 2023. MSilent: Towards General Corpus Silent Speech Recognition Using COTS MmWave Radar. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 1, Article 39 (mar 2023), 28 pages. <https://doi.org/10.1145/3580838>
- [65] Yuke Zhang, Ken Takaki, Hiroaki Murakami, Takuya Sasatani, and Yoshihiro Kawahara. 2023. Toward Continuous Finger Positioning on Ear Using Bone Conduction Speaker. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems* (Boston, Massachusetts) (*SenSys '22*). Association for Computing Machinery, New York, NY, USA, 847–848. <https://doi.org/10.1145/3560905.3568075>
- [66] Quan Zhou, Bin Fang, Jianhua Shan, Fuchun Sun, and Di Guo. 2020. A Survey of the Development of Wearable Devices. In *2020 5th International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE, Shenzhen, China, 198–203. <https://doi.org/10.1109/ICARM49381.2020.9195351>

Appendices

A GRDA APPLICATION

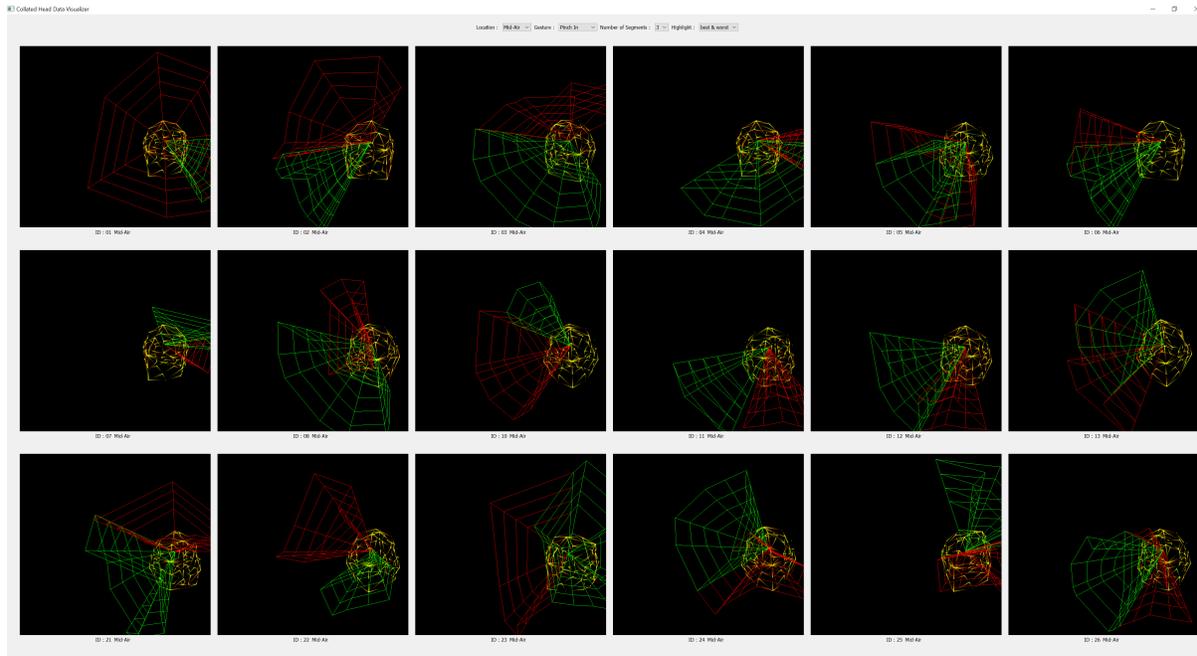


Fig. 14. GRDA application allowing observation of 3D head model and best (green) / worst (red) area comparison side by side

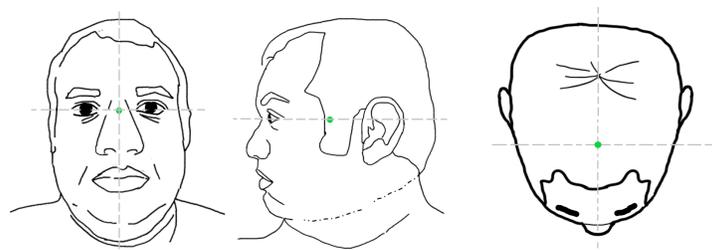


Fig. 15. Location of the center (marked in green) of the head with respect to front, side, and top view of the head for measuring distances in GRDA application. The dashed line in grey represents axis lines in 3 dimensions.

B STATISTICAL ANALYSIS

B.1 Mean and Median Values

Gesture	Gesture Time (DV1)					
	Mean			Median		
	3 Regions	5 Regions	7 Regions	3 Regions	5 Regions	7 Regions
All gestures On-skin	3712.687	3816.186	3953.525	3703.757	3819.503	3922.156
All gestures Mid-air	3536.410	3672.517	3782.245	3495.404	3593.735	3774.913
Swipe up On-skin	3493.432	3672.132	3893.627	3480.184	3691.117	3892.43
Swipe up Mid-air	3243.691	3418.072	3460.292	3304.489	3429.727	3431.065
Swipe up On-skin	3657.619	3750.83	3891.866	3660.206	3756.051	3905.35
Swipe up Mid-air	3300.891	3471.041	3713.693	3293.885	3475.328	3705.572
Swipe front On-skin	3459.888	3577.415	3719.333	3492.858	3561.259	3713.978
Swipe front Mid-air	3366.385	3509.414	3864.546	3373.323	3514.852	3688.63
Swipe back On-skin	3349.439	3523.131	3603.241	3343.3	3521.587	3587.61
Swipe back Mid-air	3394.372	3494.341	3587.924	3381.961	3504.524	3593.381
Pinch in On-skin	4195.593	4193.104	4281.091	4194.622	4171.085	4273.56
Pinch in Mid-air	4052.149	4149.953	4215.33	4036.132	4118.786	4207.033
Pinch out On-skin	4187.546	4265.589	4375.578	4177.971	4253.219	4389.343
Pinch out Mid-air	3992.793	4102.66	4110.475	3992.889	4118.463	4108.293
Tap On-skin	3650.159	3759.947	3900.269	3646.839	3798.125	3898.539
Tap Mid-air	3424.899	3549.574	3676.62	3412.574	3544.269	3682.643

Table 1. Mean and median gesture time (DV1) in milliseconds (ms).

Gesture	Gesture Path Length (DV2)					
	Mean			Median		
	3 Regions	5 Regions	7 Regions	3 Regions	5 Regions	7 Regions
All gestures On-skin	1422.108	1391.224	1409.538	1348.711	1362.364	1367.667
All gestures Mid-air	1517.834	1428.447	1445.943	1393.449	1400.107	1390.451
Swipe up On-skin	1464.479	1440.299	1402.672	1327.588	1380.54	1347.581
Swipe up Mid-air	1613.46	1509.911	1538.699	1609.951	1556.02	1528.444
Swipe up On-skin	1517.034	1418.078	1486.041	1517.034	1418.078	1486.041
Swipe up Mid-air	1703.035	1512.876	1551.334	1703.035	1512.876	1551.334
Swipe front On-skin	1465.024	1544.209	1558.211	1476.635	1484.915	1476.364
Swipe front Mid-air	1738.389	1675.337	1612.322	1739.876	1627.532	1597.57
Swipe back On-skin	1511.529	1424.764	1480.937	1472.201	1408.686	1436.66
Swipe back Mid-air	1590.3	1502.576	1550.981	1484.618	1507.982	1476.923
Pinch in On-skin	1333.953	1236.612	1229.839	1243.066	1228.667	1197.21
Pinch in Mid-air	1146.523	1160.207	1154.841	1117.967	1131.603	1093.265
Pinch out On-skin	1269.022	1237.692	1260.57	1223.834	1239.05	1242.855
Pinch out Mid-air	1277.779	1134.834	1247.689	1235.037	1108.712	1167.867
Tap On-skin	1393.713	1436.913	1448.494	1415.095	1440.751	1417.567
Tap Mid-air	1555.35	1503.385	1465.733	1413.159	1510.262	1462.826

Table 2. Mean and median gesture path length (DV2) in millimeters (mm)

Gesture	Gesture Accuracy (DV3)					
	Mean			Median		
	3 Regions	5 Regions	7 Regions	3 Regions	5 Regions	7 Regions
All gestures On-skin	95.767	91.746	81.406	100	90.0	78.571
All gestures Mid-air	94.974	84.762	78.061	100	90.0	78.571
Swipe up On-skin	99.074	95.556	82.54	100	100	85.714
Swipe up Mid-air	93.518	81.667	78.968	100	80	78.571
Swipe up On-skin	96.296	93.889	83.333	100	90	85.714
Swipe up Mid-air	97.222	88.333	77.381	100	90	78.571
Swipe front On-skin	97.222	94.444	83.73	100	90	85.714
Swipe front Mid-air	97.222	86.111	75.793	100	90	71.429
Swipe back On-skin	89.815	82.778	78.968	83.333	85	75.0
Swipe back Mid-air	96.296	87.778	86.905	100	90	85.714
Pinch in On-skin	98.148	91.667	80.158	100	90	78.571
Pinch in Mid-air	91.667	82.775	76.19	91.667	80	71.429
Pinch out On-skin	91.667	90.556	78.968	91.667	90	78.871
Pinch out Mid-air	92.593	80.556	78.147	100	80	78.571
Tap On-skin	98.148	93.333	82.143	100	90	78.571
Tap Mid-air	96.296	86.111	73.016	100	90	71.429

Table 3. Mean and median gesture accuracy (DV3) in percentage (%).

B.2 Gesture Time (DV1) Analysis

Gesture	# of Gesture Regions in Interaction Space	Gesture Time (DV1)							
		On-skin Vs Mid-air Gestures							
		Shapiro-Wilk Test Results		Bartlett Test Results		Unpaired T-test Results for Normally Distributed Data		Mann-Whitney U Test Results for Data with Non-Normal Distribution	
	<i>W</i>	<i>p</i>	$\chi^2(1)$	<i>p</i>	<i>t</i>	<i>p</i>	<i>W</i>	<i>p</i>	
All gestures	3	0.994	< 0.001	-	-	-	-	220569	< 0.001
	5	0.989	< 0.001	-	-	-	-	614211	< 0.001
	7	0.998	< 0.001	-	-	-	-	1195028	< 0.001
Swipe up	3	0.993	0.423	0.216	0.642	$t(213.57) = 5.966$	< 0.001	-	-
	5	0.995	0.230	6.562	0.010	-	-	8509	< 0.001
	7	0.998	0.738	0.022	0.881	$t(501.96) = 13.514$	< 0.001	-	-
Swipe down	3	0.984	0.016	-	-	-	-	2611	< 0.001
	5	0.973	< 0.001	-	-	-	-	8544	< 0.001
	7	0.997	0.414	0.129	0.719	$t(501.74) = 7.008$	< 0.001	-	-
Swipe front	3	0.990	0.162	0.348	0.555	$t(213.3) = 2.064$	0.040	-	-
	5	0.997	0.842	37.386	< 0.001	-	-	14043	0.03
	7	0.997	0.411	0.818	0.366	$t(500.37) = 1.458$	0.146	-	-
Swipe back	3	0.995	0.641	7.973	0.005	-	-	6359	0.252
	5	0.993	0.068	35.707	< 0.001	-	-	14972	0.214
	7	0.998	0.868	21.787	< 0.001	-	-	31009	0.650
Pinch in	3	0.994	0.548	2.589	0.108	$t(208.98) = 3.703$	< 0.001	-	-
	5	0.993	0.071	0.715	0.398	$t(356.57) = 1.411$	0.159	-	-
	7	0.996	0.276	0.391	0.532	$t(501.22) = 2.423$	0.016	-	-
Pinch out	3	0.989	0.091	1.792	0.181	$t(210.49) = 4.496$	< 0.001	-	-
	5	0.997	0.629	0.703	0.402	$t(356.6) = 5.021$	< 0.001	-	-
	7	0.997	0.601	0.337	0.562	$t(501.33) = 10.105$	< 0.001	-	-
Tap	3	0.995	0.756	3.408	0.065	$t(207.47) = 4.969$	< 0.001	-	-
	5	0.994	0.179	36.576	< 0.001	-	-	8214	< 0.001
	7	0.994	0.022	31.425	< 0.001	-	-	17018	< 0.001

Table 4. Statistical Test Report for analyzing the effect of interaction space choice on gesture time (DV1) for different number of gesture (Swipe/Pinch/Tap) regions around the ear. Yellow boxes represent statistical insignificance.

Gesture	Gesture Time (DV1)										
	Shapiro-Wilk Test Results		Bartlett-Test Results		One-Way ANOVA Results for Normally Distributed Data		Kruskal-Wallis Test Results for Data with Non-Normal Distribution		Post-hoc Test Results		
	<i>W</i>	<i>p</i>	$\chi^2(2)$	<i>p</i>	<i>F</i> (2, 502)	<i>p</i>	$\chi^2(2)$	<i>p</i>	3 Region vs 5 Region	5 Region vs 7 Region	3 Region vs 7 Region
All gestures On-skin	0.999	0.097	54.765	< 0.001	-	-	157.96	< 0.001	< 0.001	< 0.001	< 0.001
All gestures Mid-air	0.994	< 0.001	-	-	-	-	210.23	< 0.001	< 0.001	< 0.001	< 0.001
Swipe up On-skin	0.992	0.004	-	-	-	-	102.32	< 0.001	< 0.001	< 0.001	< 0.001
Swipe up Mid-air	0.997	0.570	9.657	0.008	-	-	27.286	< 0.001	< 0.001	0.72	< 0.001
Swipe down On-skin	0.987	< 0.001	-	-	-	-	27.458	< 0.001	0.252	< 0.001	< 0.001
Swipe down Mid-air	0.995	0.051	48.396	< 0.001	-	-	165.79	< 0.001	< 0.001	< 0.001	< 0.001
Swipe front On-skin	0.992	0.007	-	-	-	-	48.043	< 0.001	0.041	< 0.001	< 0.001
Swipe front Mid-air	0.993	0.009	-	-	-	-	100.79	< 0.001	< 0.001	< 0.001	< 0.001
Swipe back On-skin	0.991	0.003	-	-	-	-	42.3	< 0.001	< 0.001	0.019	< 0.001
Swipe back Mid-air	0.998	0.647	36.365	< 0.001	-	-	31.522	< 0.001	0.010	0.002	< 0.001
Pinch in On-skin	0.996	0.289	0.286	0.867	8.094	< 0.001	-	-	1.0	0.010	0.046
Pinch in Mid-air	0.995	0.060	2.658	0.265	20.701	< 0.001	-	-	0.015	< 0.001	0.071
Pinch out On-skin	0.998	0.672	2.725	0.256	42.535	< 0.001	-	-	0.112	< 0.001	< 0.001
Pinch out Mid-air	0.998	0.913	1.864	0.394	57.640	< 0.001	-	-	0.008	1.0	0.002
Tap On-skin	0.974	< 0.001	-	-	-	-	50.330	< 0.001	0.025	< 0.001	< 0.001
Tap Mid-air	0.996	0.117	47.823	< 0.001	-	-	50.277	< 0.001	< 0.001	< 0.001	< 0.001

Table 5. Statistical Test Report for analyzing the effect of increasing the number of regions around the ear on gesture time (DV1) for off-device mid-air and on-skin gesture (Swipe/Pinch/Tap) reuse. Yellow boxes represent statistical insignificance.

B.3 Gesture Path Length (DV2) Analysis

Gesture	# of Gesture Regions in Interaction Space	Gesture Path Length (DV2)							
		Onskin Vs Mid-air Gestures							
		Shapiro-Wilk Test Results		Bartlett Test Results		Unpaired T-test Results for Normally Distributed Data		Mann-Whitney U Test Results for Data with Non-Normal Distribution	
		<i>W</i>	<i>p</i>	$\chi^2(1)$	<i>p</i>	<i>t</i>	<i>p</i>	<i>W</i>	<i>p</i>
All gestures	3	0.952	< 0.001	-	-	-	-	306747	0.012
	5	0.930	< 0.001	-	-	-	-	829754	0.048
	7	0.850	< 0.001	-	-	-	-	1626208	0.020
Swipe up	3	0.893	< 0.001	-	-	-	-	7150	0.004
	5	0.991	0.03	-	-	-	-	18247	0.038
	7	0.984	< 0.001	-	-	-	-	37674	< 0.001
Swipe down	3	0.957	< 0.001	-	-	-	-	6560	0.113
	5	0.995	0.267	5.906	0.015	-	-	18055	0.060
	7	0.983	< 0.001	-	-	-	-	34748	0.067
Swipe front	3	0.985	0.024	-	-	-	-	7293	< 0.001
	5	0.796	< 0.001	-	-	-	-	18469	0.021
	7	0.986	< 0.001	-	-	-	-	33820	0.206
Swipe back	3	0.961	< 0.001	-	-	-	-	6108	0.549
	5	0.991	0.048	-	-	-	-	18080	0.056
	7	0.507	< 0.001	-	-	-	-	33408	0.311
Pinch in	3	0.900	< 0.001	-	-	-	-	4262	< 0.001
	5	0.997	0.785	1.2158	0.270	$t(355.59) = 2.081$	0.038	-	-
	7	0.986	< 0.001	-	-	-	-	27424	0.008
Pinch out	3	0.962	< 0.001	-	-	-	-	5789	0.926
	5	0.985	< 0.001	-	-	-	-	12731	< 0.001
	7	0.975	< 0.001	-	-	-	-	30187	0.339
Tap	3	0.960	< 0.001	-	-	-	-	6375	0.238
	5	0.970	< 0.001	-	-	-	-	17406	0.222
	7	0.992	0.018	-	-	-	-	32853	0.501

Table 6. Statistical Test Report for analyzing the effect of interaction space choice on gesture path length (DV2) for different number of gesture (Swipe/Pinch/Tap) regions around the ear. Yellow boxes represent statistical insignificance.

Gesture	Gesture Path Length (DV2)						
	Shapiro-Wilk Test Results		Kruskal-Wallis Results for Data with Non-Normal Distribution		Post-hoc Test Results		
	<i>W</i>	<i>p</i>	$\chi^2(2)$	<i>p</i>	3 Region Vs 5 Region	5 Region Vs 7 Region	3 Region Vs 7 Region
All gestures On-skin	0.964	< 0.001	1.419	0.492	-	-	-
All gestures Mid-air	0.841	< 0.001	6.889	0.032	0.025	1.0	0.170
Swipe up On-skin	0.929	< 0.001	1.231	0.54	-	-	-
Swipe up Mid-air	0.991	0.002	2.12	0.347	-	-	-
Swipe down On-skin	0.993	0.011	2.294	0.318	-	-	-
Swipe down Mid-air	0.946	< 0.001	4.117	0.128	-	-	-
Swipe front On-skin	0.935	< 0.001	0.686	0.71	-	-	-
Swipe front Mid-air	0.891	< 0.001	5.033	0.081	-	-	-
Swipe back On-skin	0.99	< 0.001	2.654	0.265	-	-	-
Swipe back Mid-air	0.542	< 0.001	0.29	0.865	-	-	-
Pinch in On-skin	0.974	< 0.001	2.025	0.363	-	-	-
Pinch in Mid-air	0.975	< 0.001	0.357	0.836	-	-	-
Pinch out On-skin	0.984	< 0.001	0.195	0.907	-	-	-
Pinch out Mid-air	0.965	< 0.001	12.02	0.002	0.004	0.018	1.0
Tap On-skin	0.983	< 0.001	0.172	0.918	-	-	-
Tap Mid-air	0.971	< 0.001	0.388	0.824	-	-	-

Table 7. Statistical Test Report for analyzing the effect of increasing the number of regions around the ear on gesture path length (DV2) for off-device mid-air and on-skin gesture (Swipe/Pinch/Tap) reuse. Yellow boxes represent statistical insignificance.

B.4 Gesture Accuracy (DV3) Analysis

Gesture	# of Gesture Regions in Interaction Space	Gesture Accuracy (DV3)								
		Shapiro-Wilk Test Results		Bartlett Test Results		Onskin Vs Mid-air Gestures				
		W	p	$\chi^2(1)$	p	Paired T-test Results for Normally Distributed Data		Wilcoxon Test Results for Data with Non-Normal Distribution		
						t	p	W	Z	p
All gestures	3	0.560	< 0.001	-	-	-	-	470	0.884	0.461
	5	0.857	< 0.001	-	-	-	-	730.5	6.220	< 0.001
	7	0.920	< 0.001	-	-	-	-	1950.5	3.003	0.002
Swipe up	3	0.514	< 0.001	-	-	-	-	0	2.450	0.031
	5	0.839	< 0.001	-	-	-	-	0	3.645	< 0.001
	7	0.776	< 0.001	-	-	-	-	12	2.335	0.033
Swipe down	3	0.485	< 0.001	-	-	-	-	4	0.577	0.772
	5	0.775	< 0.001	-	-	-	-	13	2.516	0.020
	7	0.871	< 0.001	-	-	-	-	11	2.753	0.007
Swipe front	3	0.451	< 0.001	-	-	-	-	5	0.0	1.0
	5	0.791	< 0.001	-	-	-	-	0	3.392	< 0.001
	7	0.861	< 0.001	-	-	-	-	11	3.208	< 0.001
Swipe back	3	0.628	< 0.001	-	-	-	-	40	2.333	0.039
	5	0.827	< 0.001	-	-	-	-	100	1.616	0.152
	7	0.830	< 0.001	-	-	-	-	135	2.784	0.003
Pinch in	3	0.580	< 0.001	-	-	-	-	5	2.333	0.039
	5	0.870	< 0.001	-	-	-	-	8	2.673	0.008
	7	0.917	< 0.001	-	-	-	-	40	0.729	0.481
Pinch out	3	0.636	< 0.001	-	-	-	-	25	0.333	0.789
	5	0.879	< 0.001	-	-	-	-	11	2.922	0.004
	7	0.909	0.006	-	-	-	-	69	0.044	0.961
Tap	3	0.451	< 0.001	-	-	-	-	7	0.817	0.688
	5	0.806	< 0.001	-	-	-	-	0	3.118	0.002
	7	0.915	0.009	-	-	-	-	10	3.057	0.002

Table 8. Statistical Test Report for analyzing the effect of interaction space choice on gesture accuracy (DV3) for different numbers of gesture (Swipe/Pinch/Tap) regions around the ear. Yellow boxes represent statistical insignificance.

Gesture	Gesture Accuracy (DV3)						
	Shapiro-Wilk Test Results		Friedman Test Results for Data with Non-Normal Distribution		Post-hoc Test Results		
	W	p	$\chi^2(2)$	p	3 Region Vs 5 Region	5 Region Vs 7 Region	3 Region Vs 7 Region
Swipe up On-skin	0.874	< 0.001	117.05	< 0.001	< 0.001	< 0.001	< 0.001
Swipe up Mid-air	0.920	< 0.001	129.27	< 0.001	< 0.001	< 0.001	< 0.001
Swipe up On-skin	0.777	< 0.001	23.311	< 0.001	0.4281	0.003	< 0.001
Swipe up Mid-air	0.846	< 0.001	20.333	< 0.001	0.002	0.115	< 0.001
Swipe down On-skin	0.867	< 0.001	23.311	< 0.001	0.4281	0.003	< 0.001
Swipe down Mid-air	0.877	< 0.001	20.333	< 0.001	0.003	0.115	< 0.001
Swipe front On-skin	0.827	< 0.001	24.125	< 0.001	0.077	< 0.001	< 0.001
Swipe front Mid-air	0.876	< 0.001	32.444	< 0.001	< 0.001	0.002	< 0.001
Swipe back On-skin	0.921	< 0.001	5.778	0.045	0.033	0.180	0.008
Swipe back Mid-air	0.871	< 0.001	10.111	0.006	0.027	1	0.002
Pinch in On-skin	0.847	< 0.001	24.121	< 0.001	0.032	0.007	< 0.001
Pinch in Mid-air	0.923	0.002	12.333	0.002	0.133	0.025	0.003
Pinch out On-skin	0.901	< 0.001	6.971	0.031	1	0.003	0.009
Pinch out Mid-air	0.928	0.003	10.333	0.006	0.017	1	0.004
Tap On-skin	0.831	< 0.001	19.5	< 0.001	0.036	0.002	< 0.001
Tap Mid-air	0.905	< 0.001	29.778	< 0.001	0.003	< 0.001	< 0.001

Table 9. Statistical Test Report for analyzing the effect of increasing the number of regions around the ear on gesture accuracy (DV3) for off-device mid-air and on-skin gestures (Swipe/Pinch/Tap). Yellow boxes represent statistical insignificance.