

Exploring Above-Neck Unimanual Swipe Gestures for Off-Device Earable Interaction

Shaikh Shawon Arefin Shimon

ssarefin@uwaterloo.ca
University of Waterloo
Waterloo, Ontario, Canada

Ali Neshati

ali.neshati@ontariotechu.ca
Ontario Tech University
Oshawa, Ontario, Canada

Junwei Sun

jwsun@idt.eitech.edu.cn
Ningbo Institute of Digital Twin
Ningbo, China

Qiang Xu

qiang.xu1@huawei.com
Huawei Technologies Canada Co., Ltd.
Markham, Ontario, Canada

Jian Zhao

jianzhao@uwaterloo.ca
University of Waterloo
Waterloo, Ontario, Canada

Abstract

Despite their growing popularity, in-ear *Earable* / *Hearable* devices (i.e., ear-mounted wearables) face interaction challenges due to limited input space and compact form factors. To enhance interaction capabilities, researchers are exploring off-device hand-based input spaces above the neck using midair and onskin gestures. However, existing literature primarily focuses on axial swipes (i.e., horizontal and vertical), leaving nonaxial swipes (i.e., unidirectional swipes with varied orientations) and angular swipes (e.g., “L”, “U”, or “V”) largely underexplored despite their potential interaction advantages. To address this gap, we conducted a within-subject gesture motion analysis study with 24 participants, analyzing 5,568 swipes of varying shape, orientation, and complexity. Our results revealed preferred starting and ending regions for different unidirectional and angular swipe shapes, as well as intuitive swipe shapes within the off-device, above-neck manual interaction space. We further examine off-device swipe characteristics, discuss the feasibility of recognizing these earable gestures with current sensing technologies, and highlight their potential application in various scenarios. These findings broaden the understanding of off-device earable gestures and provide design insights for integrating suitable nonaxial and angular swipes alongside traditional axial gestures to enhance interaction with in-ear earable devices.

CCS Concepts

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

Keywords

Embodied Interaction, Input Techniques, Unimanual Interaction, Hand-to-face Gestures, Earables

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

GI '26, Kitchener - Waterloo, Ontario, Canada

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-XXXX-X/2026/06

<https://doi.org/XXXXXXXX.XXXXXXX>

ACM Reference Format:

Shaikh Shawon Arefin Shimon, Ali Neshati, Junwei Sun, Qiang Xu, and Jian Zhao. 2026. Exploring Above-Neck Unimanual Swipe Gestures for Off-Device Earable Interaction. In *Proceedings of Graphics Interface 2026 (GI '26)*. ACM, New York, NY, USA, 20 pages. <https://doi.org/XXXXXXXX.XXXXXXX>

1 Introduction

Ear-mounted wearables (i.e., *earables* / *hearables*) are widely used for their auditory capabilities, and are increasingly explored as head-based sensing platforms [67]. Wireless in-ear devices, such as *Apple AirPods* [5], are especially popular for their compactness, portability, and convenience [98], but their small form factor imposes substantial constraints on direct on-device interaction. Simple actions such as *accepting or rejecting calls* and *adjusting volume* can be performed through taps, multi-presses, or vertical swipes on the device body, more complex tasks (e.g., *menu navigation* or *application switching*) are challenging due to limited on-device input space. While *Voice* input can partially mitigate these constraints, it is often less effective than other input methods for head-mounted devices [45], may be socially inappropriate in certain contexts, [60], and is susceptible to environmental noise [49]. These limitations motivate exploring alternative, off-device interaction techniques for in-ear earables.

Researchers have explored various off-device modalities – such as midair and onskin manual gestures, head motion, full-body movement, and silent commands – to extend in-ear earable interaction beyond physical, on-device input [2, 39, 51, 90]. Among these, unimanual (i.e., one-handed) hand-to-face gestures above the neck have emerged as particularly appealing to end-users [67]. These gestures occur naturally in daily behavior [46, 57], and facilitate one-handed interaction with the ear-mounted device while keeping the other hand available for additional tasks. Prior studies also show that swipe gestures performed in midair and onskin regions above the neck are among the most preferred unimanual off-device gestures for earables [13, 64]. However, direct physical interaction with the outer ear can cause in-ear earables to displace from the ear canal [36], making direct touch interactions on the ear itself unsuitable for interacting with these devices. This motivates a focus on unimanual hand-to-face swipe gestures performed in the onskin and midair spaces *around* and *above* the outer ear, rather than on the ear surface itself.

Previous work has demonstrated that unimanual gestures can be differentiated by both shape and location in midair and onskin

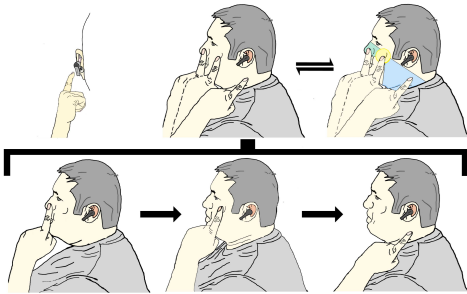


Figure 1: Unimanual (one-handed) swipe using the left hand from a starting (Green) region S through a swipe focal point (Yellow) to an ending (Blue) region E on the left hand side of the face.

interaction spaces around the ear [3, 43, 51, 90]. More recently, Shimon et al. [6] explored location-based reuse of axial (horizontal and vertical) swipe, tap, and pinch gestures for off-device earable input. Despite these advances, most prior studies have focused primarily on axial swipe directions [27, 67, 90]. While swipe or stroke gestures of varying complexity, difficult ranking and number of strokes in the above-neck space have been proposed [79], there is limited understanding of how nonaxial unidirectional swipes and angular swipe shapes – such as “L”, “U”, and “V” – perform in off-device earable interaction. This gap is notable, as previous studies highlight that these non-axial and angular swipes yield above-average user agreement [64]. These unistroke gestures also exhibit low shape complexity and low execution difficulty [37, 79, 80]. Prior work on symbolic and marking menu-based interaction further demonstrates that users naturally employ these swipes resembling checkmark (✓) or caret (Λ) for semantically meaningful actions, (e.g., confirmation, selection) across various input modalities, including onskin and midair interactions [79, 85, 87]. By allowing angular swipe / stroke gestures with fixed contortions to be rotated or re-oriented, we can expand the off-device input vocabulary and encode distinct earable input commands while avoiding precision issues associated with small gesture region boundaries suggested in location-based gesture reuse [6] for midair and onskin, off-device earable spaces. Furthermore, these expressive, angular manual gestures can serve as delimiters to mitigate the *Midas Touch* problem in manual hand-to-face interaction for head-based wearables [84] as their explicit and deliberate articulation minimizes the chance of unintentional gesture activation.

To address this research gap, we conducted a controlled within-subject study with 24 participants, analyzing 5,568 unimanual hand-to-face swipes performed both in midair above the ear, and on the skin around the ear for off-device interaction with an in-ear earable. Rather than being limited by current gesture-sensing constraints of such devices, we adopted an *imaginary interface* approach to examine user-preferred motion independent of sensing limitations [17, 28], and employed an optical motion-capture system [6, 14, 22] for detailed analysis of unidirectional and angular single-stroke swipe gestures. We examined axial (horizontal and vertical), non-axial (unidirectional but non-horizontal / vertical), and angular (e.g., “L”, “U”, and “V”) swipes that start at a defined region, pass through a focal point near a facial landmark, and end at another region

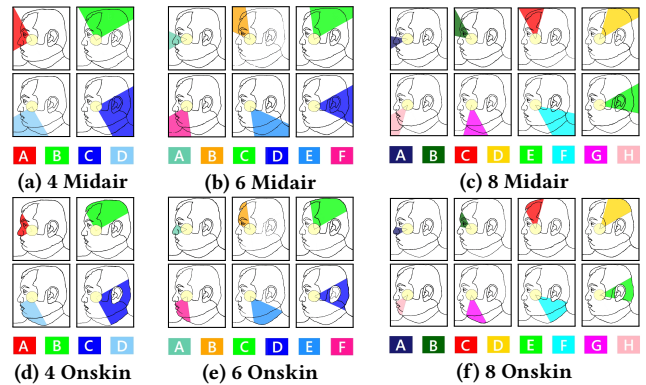


Figure 2: 4, 6, and 8 region segmentation on (onskin) / above (midair) a swipe focal point (e.g., cheekbone, marked in Yellow).

(Figure 1). We systematically studied spatial density effects on various swipe shapes by evaluating region segmentations of varying granularity (4, 6, and 8 regions) centered around the cheekbone focal region (Figure 2).

Our findings show that although participants preferred onskin swipes at lower region densities, unidirectional and angular midair swipes scaled more effectively as swipe region density increased, outperforming onskin interaction in speed and precision under highly segmented conditions. However, across both interaction spaces, usability degraded beyond the 6-region segmentation, with several onskin cheekbone regions exhibiting poor usability even in the 6-region condition.

Taken together, these findings inform our proposed design guidelines for structuring practical subsets of unidirectional and angular swipe gestures for off-device earable interaction, which can enable more intuitive and efficient interaction within increasingly segmented above-neck input spaces. By characterizing the usability and scalability limits of midair and onskin swipe regions, our contribution extends prior research on unimanual off-device earable interaction and provides a foundation for richer, more scalable gesture vocabulary design for future in-ear earable systems.

2 Related Work

2.1 Off-Device Earable Interaction Techniques

Earable interaction researchers have proposed *facial expressions* [15, 71, 82, 95], *silent commands* [18, 39, 55, 94], *jaw motion-based gestures* [8, 42, 72, 73], *head movement* [24, 25, 34, 63] and *body movement-based gestures* [99] as potential non-manual approaches for off-device interaction with earable devices. However, facial expression-based gestures have less acceptability in public settings [44] due to the possibility of attracting unwarranted, negative attention [20]. 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 [23] in other wearable or interactive devices. After repeated usage, such gestures can generate jaw and facial muscle fatigue among end-users. Large whole-body movement-based interactions have been observed to attract negative attention in public spaces [13] similar to approaches based on facial expression. Although proposed head-movement interaction

models are more subtle than whole-body interaction, whole-body and head-movement-based gestures can evoke fatigue and physical strain to gesture frequency and path length.

Prior work demonstrates that manual (i.e., hand-based) gestures are often unconscious and spontaneous behaviors [7, 21, 30, 56], with hand-to-face motions (e.g., face touching) occurring as frequently as 15.7 to 23 times per hour [57]. In natural, unconstrained settings, users have shown to prefer subtle above-neck unimanual gestures rather than larger full-body movements, especially when at least one hand is free [33]. These gestures are also more socially acceptable, as large manual motions attract unwanted attention [44]. Above-neck gestures align with user expectations for in-ear earables, where one hand interacts with a single earbud, leaving the alternate hand available for other possible tasks. Hand-to-face onskin or midair gestures also avoid the *Gorilla Arm* fatigue caused by large midair movements away from the body [32]. These benefits make unimanual, hand-to-face gestures a popular choice for off-device earable input [10, 16, 67].

Earable systems capable of recognizing midair and onskin unimanual gestures have been proposed in *FreeDigiter* [53], *SonicASL* [40], *Earbuddy* [90], *FaceSense* [41], *OESense* [52], *EarGest* [2], *EarAce* [10], *LeakyFeeder* [92], *EarHover* [74], *MAF* [93], *UiEar* [97], *TinyssimoRadar* [68]. These works (revisited in detail in Section 6.2) show that it is viable to detect both gesture shape and gesture location for midair and onskin spaces above the neck for unimanual swipes, which sets the foundation of our study.

2.2 Input Space Segmentation for Enabling Gesture Variations

While complex gesture shapes and recognition techniques have been widely studied across wearable devices, interaction-space segmentation to enable gesture variation through location, orientation, and contortion remains underexplored for earables. Smartwatch research has examined dividing circular bezels into multiple regions to support location-based swipe variation [66, 88], while Dezfuli et al. [17] segmented an imaginary hand / palm gesture space to enable onskin swipe and tap input. Li et al. [50] explored pie-menus for midair gesture interaction with remote displays using front-of-body manual gestures. Similarly, Tu et al. [75] explored radial menu interaction for distance displays with lateral arm movements in front of the body. Gil et al. [26] explored region placement and region count for in-air thumb typing on a commercial head-mounted display (HMD). Although these works target other wearables, their evaluation metrics and study designs inform gesture reuse across multiple regions in alternative interaction spaces.

Lissermann et al. [51] studied region-based reuse of unimanual touches, taps, and grasp on and above the outer ear for earable interaction, but did not examine continuous swipes or gestures on regions beyond the outer ear. Zhang et al. [96] examined onskin tap recognition across 5 outer-ear regions using a bone-conduction earable device. Elicitation studies by Hanae et al. [64] and Chen et al. [13] show that users prefer subtle unimanual gestures in onskin and midair areas above the neck, with single-finger horizontal / vertical swipes, taps, and two-finger pinches receiving the highest acceptance. Building on this work, Shimon et al. [6] extended outer-ear segmentation to midair and onskin spaces around

the ear using non-overlapping regions, identifying key boundaries for the highest-agreement gesture primitives reported in [64].

However, the location-based swipe boundaries proposed in [6] constrain natural swipe speed and gesture extent in order to avoid region boundary crossing. While prior earable work has largely focused on axial swipes, elicitation study results from [64] demonstrate above-average acceptance of non-axial and angular swipes. Varying swipe orientation and angular contortion can therefore expand the earable input vocabulary in both midair and onskin spaces without relying on densely segmented, non-overlapping regions. Although non-axial and angular swipes have been studied on smartwatch screens and bezels [66, 88], comparable analyses for hand-to-face midair and onskin swipes in off-device earable interaction remain unavailable, which we address in this paper.

2.3 Gesture Analysis using Motion Capture Data

Beyond generating ground-truth data for on-device gesture recognition systems in development, motion-capture technologies such as *Vicon* [76], *OptiTrack* [59], and *Kinect* [54] are widely used to study gesture motion and showcase proof-of-concepts for novel interaction techniques [62, 83]. By decoupling interaction design from sensing limitations, this imaginary-interface approach enables early investigation of emerging interaction spaces before robust on-device sensing support matures in mobile and wearable devices.

In the context of VR interaction, Faleel et al. [22] explored *Hand Proximate User Interfaces* (HPUIs) for head-mounted displays using *Vicon* tracking when HMD hand tracking remained unreliable under self-occlusion, while Holfeld et al. [61] later used *Vicon* motion tracking to evaluate HPUI design guidelines independent of headset sensing limitations. Similarly, *BodyTrack* by Cheng et al. [14] used an *OptiTrack*-based motion capture system to explore full-body interaction spaces and identify potential gestures for VR headsets before integrated full-body tracking hardware became widely available.

Beyond VR systems, Dezfuli et al. [17] used *OptiTrack*-based motion tracking to explore an imaginary palm-based interactive surface for TV control. Shimon et al. [6] investigated location-based gesture reuse for unimanual off-device earable interaction using a *Vicon* system instead of earable hardware, motivated by advances in above-neck gesture sensing for ear-mounted wearables. Inspired by and grounded in these proof-of-concept approaches, we leverage the high-precision tracking capability of *Vicon*-based motion capture system in this study to extensively investigate off-device earable swipes in both onskin and midair input spaces. We believe our collected data and analysis could be valuable for interaction space exploration to support the development of capable gesture recognition systems.

3 Experiment

3.1 Experiment Overview

This study examines unidirectional and angular hand-to-face swipe gestures for off-device earable interaction using a marker-based motion capture system. We focus exclusively on swipe gestures, excluding other unimanual inputs such as taps and pinches due to their distinct motion characteristics and behavioral patterns [91]. Gestures were performed in above-neck midair and onskin

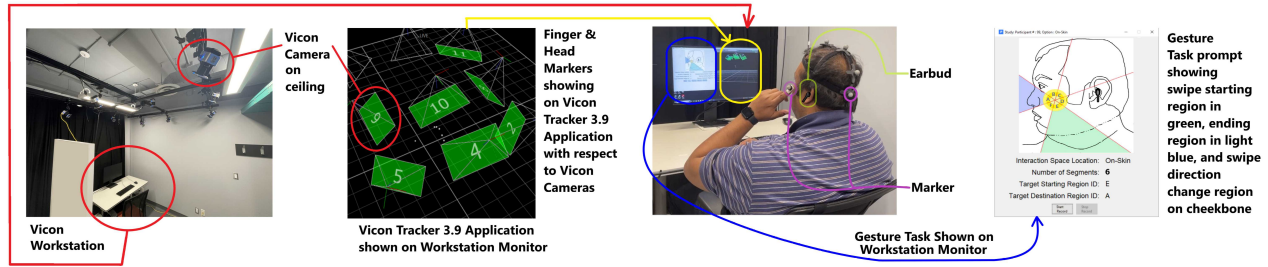


Figure 3: Different parts of study area and apparatus.

interaction spaces and analyzed within a head-centered coordinate frame to capture spatial articulation, performance, and workload characteristics. The study adopts an observational, user-centered approach to understand how users naturally articulate hand-to-face manual swipe gestures for off-device interaction with in-ear earable devices; and to inform the design of robust and usable gesture sets for such use-case.

3.2 Research Questions

This work investigates how interaction space, region density, and spatial region location influence the articulation and usability of above-neck hand-to-face swipe gestures for off-device interaction with in-ear earables, motivating the following research questions:

- **RQ1:** How do swipe performance and perceived workload differ between midair and onskin interaction spaces when the number of gesture regions is held constant?
- **RQ2:** How does increasing the number of available starting and ending regions affect swipe performance and perceived workload within a given interaction space?
- **RQ3:** Within a fixed gesture block (i.e., fixed interaction space and region density), how do swipe performance and subjective region preference vary across different starting and ending region locations?

Section 4.1 explains the metrics used to examine these research questions, and the rationale behind the metric choices.

3.3 Study Design

We employed a within-subject 2×3 factorial design with interaction space (midair vs. onskin) and region density (4, 6, and 8 regions) as independent variables. Each participant completed one gesture block for each factor combination, yielding 6 blocks in total. For each interaction space, participants performed 116 unique region-to-region swipe conditions, covering all possible starting and ending region pairings across region densities ($4 \times 4 = 16$, $6 \times 6 = 36$, $8 \times 8 = 64$). To control for order effects, the study was divided into two counterbalanced halves by interaction space, each comprising three gesture blocks, with swipe conditions presented in randomized order within each block.

3.4 Participants

We recruited 24 participants (15 male, 9 female; age 18–45, $M = 29.33$, $SD = 7.20$) from the university and surrounding community. All participants had at least a high school diploma, with 22 (90%) holding a bachelor's degree or higher, including 11 (45.8%)

with a graduate degree. All participants used their non-dominant hand for gesture interaction; 21 (87.5%) were right-handed, and 3 (12.5%) were left-handed. Although 12 participants (50%) reported regularly wearing eyewear, 21 participants completed the study without eyewear. 18 participants (75%) owned wireless earbuds (0–20 hours / week of use, $M = 11.58$, $SD = 8.37$). None reported any hand, arm, or finger joint issues that could affect gesture performance.

This study, approved by the *Institutional Review Board (IRB)*, was conducted over two 90-minute sessions on separate days to reduce fatigue, with participants receiving \$50 compensation. In total, 5,568 gesture motions were collected across all participants, interaction spaces, and swipe conditions.

3.5 Apparatus and Sensing Setup

The study used a ceiling-mounted 3×3 *Vicon* motion-capture camera grid connected to a workstation (16-core AMD Ryzen 9 CPU, 32 GB RAM) running Windows 10 Pro (Figure 3). The workstation ran *Vicon Tracker 3.9* software alongside a custom Python/PyQT-based gesture analysis application (Figure 4), with task instructions displayed on a 24-inch UHD monitor (Figure 3). Participants were seated in front of the workstation, and fitted with a head strap holding five *Vicon* markers along the sagittal plane [81], a marker on the index finger of the non-dominant hand, and a disconnected in-ear wireless earbud [58] on the non-dominant side as a prop for imagined earable-based gesture sensing. Swipe gestures were performed unimanually with the non-dominant index finger, while the dominant hand controlled a keyboard and mouse, reflecting common non-dominant hand use for secondary input [6, 19, 47]. A marked resting position on the table indicated the start and end points of each trial.

Marker trajectories were analyzed in a head-centered coordinate frame defined by the intersection of the sagittal plane and an eye-level horizontal plane, a reference frame used in analyze hand-to-face and near-head gestures [6]. Analysis was limited to motion within a 30 cm-radius spherical volume centered on the head, corresponding to the close and mid distance ranges of around-device gesture spaces [1] and capturing the spatial extent of above-neck manual interaction [6, 64]. Motion outside this region was excluded to reduce confounding from large arm movements and to ensure consistent measurement of off-device gesture timing.

3.6 Gesture Layout and Swipe Focal Point

The above-neck midair and onskin off-device interaction spaces were segmented into fan-shaped regions arranged around a shared

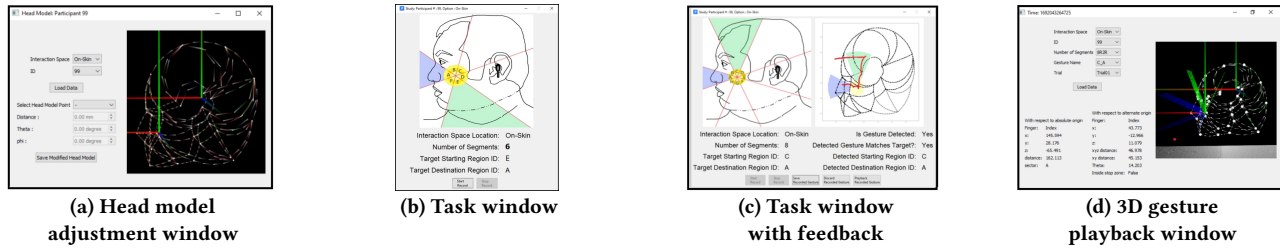


Figure 4: Different views in custom gesture analysis application.

swipe focal point (Figure 2), defined as a facial landmark at which swipe direction transitions occur. Swipe gestures were modeled as rapid unimanual index-finger motions originating in a starting region, passing through the focal point, and terminating in an ending region (Figure 1). The swipe layout was informed by prior hand-to-face interaction literature and a co-design session with (6) participants. A 4-region configuration was first adopted to support horizontal (anterior–posterior) and vertical (inferior–superior) swipes, identified as fundamental hand-to-face swipe primitives in prior work [70]. Denser configurations with 6-regions and 8-regions were subsequently evaluated to enable angular (rotational) swipes and finer-grained input.

Prior research highlights a directional bias toward horizontal panning along the posterior–anterior axis between the ear and nose in face-centric gesture spaces [70]. Building on these findings and related work on hand-to-face and earable interactions [6, 69, 90], we identified the cheekbone as the optimal focal point for onskin swipes. Its high haptic salience, visual accessibility for tracking direction changes, and balanced spacing relative to facial landmarks—such as the eye corner, nostril, nasal bridge, and jawline—support the creation of evenly distributed region boundaries. In contrast, placing the focal point above the ear for midair swipes increased muscle strain during non-axial movements, making it unsuitable for angular swipes (e.g., “L”, “U”, and “V” shapes). Therefore, a shared focal region above the cheekbone was adopted for both onskin and midair gestures.

Guided by these observations, swipe regions were oriented along the posterior–anterior axis defined between the nostril and the top of the outer ear helix for all region densities. To maintain swipe reliability, region density was capped at eight regions to avoid spans that were too narrow.

3.7 Tasks

Participants completed a series of single unimanual, unistroke swipe tasks between specified start and end regions around the cheekbone (Figure 4b). Tasks included both unidirectional straight swipes and angular swipes requiring a directional change at the focal point. Each trial began and ended with the hand at the designated resting position on the workstation: the finger entered the head-centered gesture space (Section 3.5), passed through the focal point, and returned to rest. Across gesture blocks, all valid region pairings were systematically tested.

3.8 Procedure

Each session followed a fixed procedure for administering the unidirectional and angular, above-neck swipe tasks described above.

The study was conducted over two 90-minute sessions on separate days, each corresponding to one interaction space. The first session began with a study briefing, informed consent, and a demographic survey. Each session then proceeded to a task instruction phase, and an environment walkthrough specific to that session’s interaction space. Participants were instructed to avoid touching the in-ear earable, or the outer ear to prevent the ear-mounted device from being displaced.

Before data collection, participants completed a calibration phase to generate an individualized head model for swipe analysis, with optional adjustment to ensure accurate alignment with each participant’s facial features (Figure 4a). Participants then completed counterbalanced gesture blocks presented by the custom application (Figure 4b). Each trial began and ended with the non-dominant hand resting at a designated table location, while recording was initiated and terminated using the dominant hand.

Immediate visual feedback of the recorded swipe trajectory was presented after each trial (Figures 4c and 4d), allowing participants to accept the trial, or initiate a retrieval. Participants were instructed to base retrieval decisions exclusively on the geometric correspondence between the trajectory and the individualized head model, rather than on any system-level gesture classification or recognition output. After each trial ended, feedback was displayed as a gesture trajectory visualization; however, no performance-related metrics were provided. This feedback mechanism functioned as a data quality control step, enabling detection and correction of occasional head-model misalignment or tracking artifacts. As a result, sensing- and calibration-related confounds were reduced while the influence on gesture articulation was minimized.

Trials lasted approximately 15–25 seconds, with short breaks between gesture blocks and additional rest permitted as needed. At the end of each gesture block, participants completed a brief questionnaire and participated in a short Q&A session. Each session concluded with a semi-structured interview focusing on swipe starting and ending regions, swipe shape characteristics, and overall gesture preferences. Following data collection, swipe trajectories were labeled using an integrated baseline recognizer within our custom gesture analysis application, informed by *Vicon* tracking data, with labels subsequently verified through manual review prior to accuracy analysis.

4 Measures and Analysis

4.1 Measures

Following prior HCI research on swipe and stroke gesture performance, we evaluated gesture behavior using a combination of

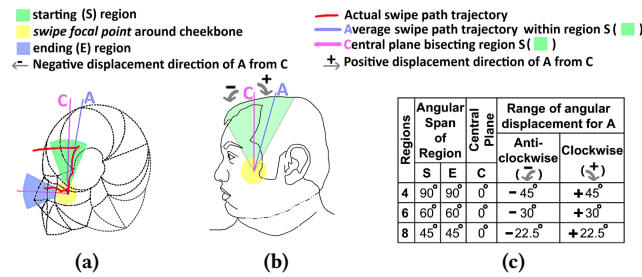


Figure 5: Swipe interaction regions and angular displacement on the left side of the face. (a) A unimanual swipe trajectory (red line) passing through the starting region (Green), focal region (Yellow), and ending region (Blue). (b) Central plane (C) and average trajectory (A) of the swipe segment within the starting region, with angular displacement defined relative to the central plane. (c) Region configurations for 4-, 6-, and 8-region distributions showing the angular span of each region and the corresponding range of angular displacement of A from the central plane C of the starting or ending region. Greater angular displacement of A from C increases the likelihood that a swipe crosses into an incorrect starting or ending region.

objective performance metrics and subjective workload and preference measures. We adopted *accuracy* [4, 77, 87], *gesture articulation time* [9, 29, 65, 87], and *gesture path length* [48] as core performance metrics, complemented by perceived workload (NASA-TLX) [31, 90] and self-reported region preferences [80, 86]. To capture spatial deviations in swipe execution, we additionally introduced *angular displacement*, inspired by the *bending error (BE)* metric for stroke gestures proposed by Vatavu et al. [78], which quantifies the magnitude and direction of swipe skew relative to region geometry.

- **DV1 – Accuracy:** Percentage (%) of gestures correctly initiated and terminated within the intended starting and ending regions, defined onskin or in midair.
- **DV2 – Gesture time:** Total duration (ms) from the hand entering the head-centered gesture space (Section 3.5) to completing the swipe and exiting the space.
- **DV3 – Gesture path length:** Cumulative spatial length (mm) of the swipe trajectory traced by the index finger during gesture time (DV2).
- **DV4 – Angular displacement:** Angular deviation (°) of the swipe trajectory (i.e., drift) within the starting (S) or ending (E) region relative to the region’s central reference plane (C) (Figure 5).

After each gesture block, perceived task workload was assessed using the NASA Task Load Index (NASA-TLX) on a 1–20 scale. Participants also reported preferences for swipe starting and ending regions using a single region-level rating (e.g., 1–4 for the 4-region segmentation), with higher values indicating stronger preference. RQ1 and RQ2 were evaluated using performance measures (DV1–DV4) together with NASA-TLX workload scores, while RQ3 examined region-level effects using accuracy (DV1), angular displacement (DV4), and subjective region ratings. Time- and distance-based measures (DV2, DV3) were excluded from RQ3 because differences in reach distance and movement direction across

regions confound direct region-level comparison under a fixed head posture.

4.2 Data Analysis

Statistical analyses employed repeated-measures techniques appropriate for the within-subject study design. Normality assumptions were assessed using the *Shapiro-Wilk* test. For two-level comparisons, paired *t*-tests were used when normality was satisfied, and Wilcoxon signed-rank tests were used otherwise. For comparisons involving three levels, *repeated-measures ANOVA* was applied to normally distributed data, while *Friedman tests* were used when normality assumptions were violated. Where applicable, post-hoc pairwise comparisons were conducted using paired *t*-tests or *Wilcoxon signed-rank tests* with *bonferroni correction* for multiple comparisons. A significance threshold of $\alpha = .05$ was used throughout.

5 Results

Section 5.1 presents quantitative analyses examining the effects of midair vs. onskin interaction space (RQ1) and region density (RQ2) on swipe performance and perceived workload. Building on these findings, Section 5.2 outlines how swipe starting and ending regions, as well as individual swipe shapes, influence swipe interaction through combined quantitative and qualitative analyses (RQ3), focusing on region layouts identified as effective in the preceding analyses. Together, these results motivate a refined subset of swipe interactions and a modified 5-region layout proposed in Section 6.1 to better support hand-to-face off-device interaction with earable devices.

5.1 Effect of Interaction Space (RQ1) and Region Density (RQ2) on Swipe Gestures

In the 4-region configuration, onskin axial and L-shaped swipes through the cheekbone area significantly outperformed midair swipe counterparts in both accuracy and user preference. Participants attributed this advantage to tactile feedback and clearer spatial constraints, which helped reduce accidental or unintentional contact with sensitive facial areas. While midair swipes enabled faster and larger movements, they lacked stabilizing feedback. However, as region density increased, individual region spans decreased for both interaction spaces. Onskin regions became increasingly constrained and placement-sensitive (e.g., near the eye corner), whereas midair swipes benefited from a larger available motion envelope. Accordingly, midair swipes became comparatively more effective at higher densities, reducing the relative benefit of tactile feedback (Figure 6). Performance metrics, workload ratings, and participant feedback all indicate that accuracy decreases and effort and frustration increase when more than 6 regions are used in either interaction space, establishing a practical upper bound on region density. We therefore recommend limiting swipe layouts to a maximum of 6 regions around the focal interaction area (Figure 2) to ensure reliable unidirectional and angular swipes.

The remainder of this subsection elaborates on these trends, with detailed statistical test results, including normality test results reported in Tables A.1 (RQ1) and A.2 (RQ2) in appendix A. Significant post hoc differences across interaction spaces and region densities have been summarized in Figure 6.

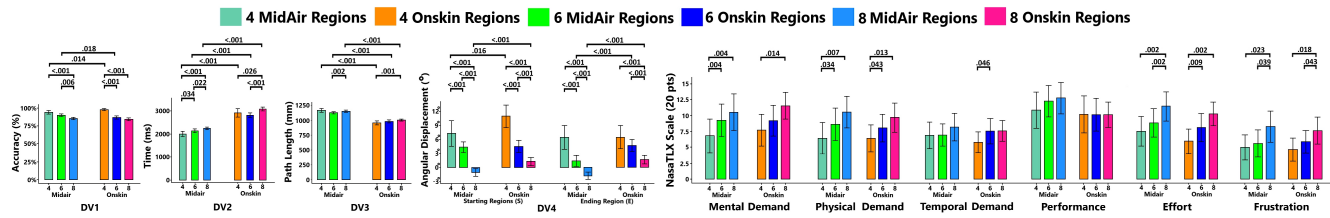


Figure 6: Gesture metrics (DV1–DV4) and NasaTLX workload for RQ1 and RQ2.

5.1.1 Gesture Metrics. Accuracy (DV1) was significantly affected by region density for both interaction spaces (Friedman test; midair $\chi^2(2) = 57.90$, onskin $\chi^2(2) = 164.76$, both $p < .001$). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction showed accuracy decreases for onskin (4 \leftrightarrow 6; $p < .001$) and midair swipes (6 \leftrightarrow 8; $p = .006$), with overall declines from 4 to 8 regions in both spaces ($p < .001$). Between interaction spaces, onskin swipes were more accurate at 4 regions (Wilcoxon $W = 148.5$, $p = .014$), midair swipes at 6 regions ($W = 5922.0$, $p = .018$), with negligible differences at 8 regions. Accuracy declined with increasing region density for both interaction spaces, with a steeper reduction for onskin swipes.

Swipe time (DV2) was significantly affected by region density for both midair (Friedman $\chi^2(2) = 55.20$, $p < .001$) and onskin swipes (Friedman $\chi^2(2) = 20.33$, $p < .001$). For onskin swipes, time increased only at higher densities (4 \leftrightarrow 8; $p < .001$; 6 \leftrightarrow 8; $p = .026$), with no significant change from 4 to 6 regions, whereas midair swipe time increased at all density steps (4 \leftrightarrow 6; $p = .034$; 6 \leftrightarrow 8; $p = .022$; 4 \leftrightarrow 8; $p < .001$). Across all region densities, onskin swipes were consistently slower than midair swipes (Wilcoxon signed-rank, all $p < .001$).

Path length (DV3) differed significantly between interaction spaces, with midair swipes producing longer paths than onskin swipes across all region densities (Wilcoxon signed-rank, all $p < .001$). Region density also had a significant effect on path length for both midair (Friedman $\chi^2(2) = 39.47$, $p < .001$) and onskin gestures (Friedman $\chi^2(2) = 16.35$, $p < .001$). Post-hoc comparisons indicated that midair path length increased only from 6 to 8 regions ($p = .002$), whereas onskin path length increased only from 4 to 8 regions ($p < .001$).

Angular displacement (DV4) varied significantly with region density for both starting and ending regions in midair and onskin interaction (Friedman test; midair start $\chi^2(2) = 148.49$, onskin start $\chi^2(2) = 148.29$, midair end $\chi^2(2) = 118.21$, onskin end $\chi^2(2) = 46.41$; all $p < .001$). Angular displacement decreased with increasing region density, with significant reductions across most comparisons ($p < .001$), except for the onskin ending region between 4 and 6 regions. At swipe initiation, midair swipes were more closely aligned with the central plane of the region than onskin swipes in the 4-region ($W = 27268$, $p = .016$) and 8-region ($W = 439705$, $p < .001$) layouts. At swipe termination, onskin swipes deviated further from the central plane of the ending region than midair swipes in the 6-region ($W = 120336$, $p < .001$) and 8-region ($W = 411172$, $p < .001$) layouts. Participants reported that skewing during both initiation and termination was manageable up to 6 regions but overly constrained in the 8-region configuration due to interference from facial features such as the eye corner.

5.1.2 NASA-TLX Perceived Workload. **Mental Demand** increased significantly with region density for both interaction spaces. For midair swipes, a significant effect was observed (Friedman $\chi^2(2) = 14.395$, $p < .001$), with increases from 4 to 6 and from 4 to 8 regions (both $p = .004$). For onskin swipes, mental demand was also significantly affected (Friedman $\chi^2(2) = 11.341$, $p < .003$), driven primarily by the increase from 4 to 8 regions ($p = .014$).

Physical Demand significantly increased with increase in region density for both midair (Friedman $\chi^2(2) = 13.300$, $p < .001$) and onskin swipes (Friedman $\chi^2(2) = 10.173$, $p < .001$). Physical demand increased from 4 to 6 regions (midair: $p = .034$; onskin: $p = .043$) and from 4 to 8 regions (midair: $p = .007$; onskin: $p = .013$), with no additional increase from 6 to 8 regions.

Temporal Demand remained largely stable across region densities, with a significant increase observed only for onskin swipes (Friedman $\chi^2(2) = 10.203$, $p < .006$) between 4 and 6 regions ($p = .046$).

Self-reported **Performance** ratings remained stable for onskin swipes, while midair performance declined with increasing region density; however, these differences were not statistically significant.

Self-reported **Effort** increased significantly with region density for both interaction spaces (Friedman; midair $\chi^2(2) = 17.82$, onskin $\chi^2(2) = 20.00$, both $p < .001$). Effort increased from 4 to 6 regions for onskin swipes ($p = .009$) and from 6 to 8 regions for midair swipes ($p = .002$), with no significant change at the other step. Overall, effort increased from 4 to 8 regions for both interaction spaces (both $p = .002$).

Similarly, **Frustration** increased significantly with region density for both midair (Friedman $\chi^2(2) = 10.265$, $p = .006$) and onskin interaction (Friedman $\chi^2(2) = 13.618$, $p = .001$). Frustration did not increase significantly from 4 to 6 regions, but increased significantly from 6 to 8 regions (midair: $p = .039$; onskin: $p = .043$) and from 4 to 8 regions (midair: $p = .023$; onskin: $p = .018$).

Overall, perceived workload remained manageable up to 6 regions, beyond which interaction became increasingly effortful and frustrating.

5.2 Effect of Starting (or Ending) Position (RQ3) Choice on Swipe Gestures

This subsection summarizes how the choice of swipe starting and ending regions influences performance in midair and onskin interaction spaces, focusing on the 4- and 6-region segmentations identified as effective in Section 5.1. Given the large volume of 4- and 6-region result, this section presents key findings for each region configuration, while detailed statistical analyses and swipe-level breakdowns are provided in Appendix B.

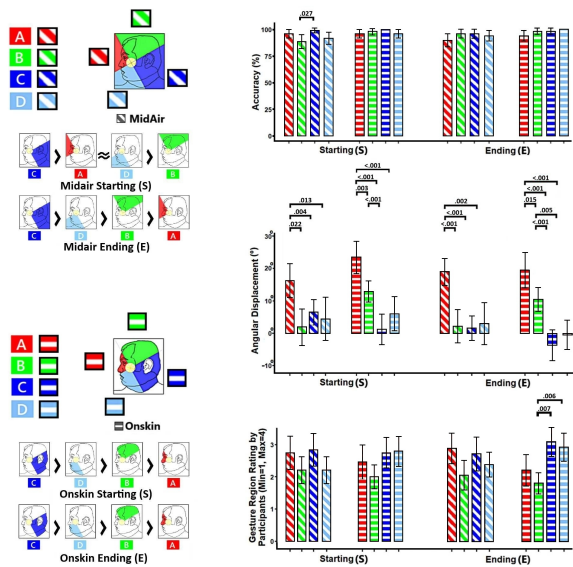


Figure 7: Gesture region layout, metrics and ranking for 4 regions.

Space	Position	Ranked Regions	Key Evidence
Midair	Start (S)	C > A ≈ D > B	C: ↑Accuracy, ↓Displacement; A: ↑displacement; good accuracy and ratings; within visual range
	End (E)	C > D > B > A	C / D: lowest displacement A: highest displacement
Onskin	Start (S)	C > D > B > A	C / D: ↓Displacement; C / D: ↑Ratings A: low rating, accuracy, high displacement
	End (E)	C > D > B > A	C / D: ↓Displacement; C / D: ↑Ratings A: low rating, accuracy, high displacement

Table 1: Relative performance of 4 swipe regions across interaction spaces. Arrows indicate metric trends contributing to the ranking (Accuracy / Ranking ↑ higher & Displacement ↓ lower is better). Accuracy = DV1, Angular Displacement = DV4, Ratings = Subjective swipe region ratings.

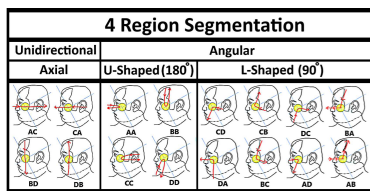


Figure 8: Visual references for individual swipe shapes in 4-region segmentation.

5.2.1 4-Region Segmentation. The analysis of the 4-region layout highlights distinct spatial interaction patterns for midair and onskin swipes above / around the cheekbone. As shown in Figure 7, the regions correspond to the anatomical contour: **A** near the nose–eye boundary, **B** at the temple, **C** around the ear and nape, and **D** along the cheek–chin area. Trends for swipe initiation and termination are summarized in Table 1, with swipe-shape stability results summarized in Tables 2a & 2b, and shape trajectories illustrated in Figure 8 for visual reference. Detailed statistical results are discussed in Appendix B.1 & B.2.

Swipe Group	Stable	Moderate	Unstable	Observation
Axial	AC, BD DB	CA	–	Axial swipes show high stability.
U-shaped (180°)	BB, CC DD	AA	–	Loops reliable near ear, temple and chin.
L-shaped (90°)	CB	BC, DC CD	AB, AD BA, DA	Angular swipes show higher variability; especially when terminating around eye.

(a) Midair

Swipe Group	Stable	Moderate	Unstable	Observation
Axial	AC, CA BD, DB	–	–	Axial swipes show high stability.
U-shaped (180°)	BB, CC DD	AA	–	U-shaped Loops reliable near ear and chin Nose loop shows larger but manageable drift.
L-shaped (90°)	–	AD, DC, CD BC, CB	AB, BA DA	Angular swipes involving region A showed weakest stability.

(b) Onskin

Table 2: Swipe shape (Figure 8) stability under 4-region segmentation derived from accuracy (DV1), angular displacement (DV4), and subjective ratings.

Both region and swipe-level analyses indicate that **axial swipes** (e.g., **AC**, **CA**, **BD**, and **DB**) exhibit consistent performance across interaction spaces (Figure 8). **U-shaped swipes** around the ear and cheek regions (**BB**, **CC**, and **DD**) demonstrate strong stability and user preference. Conversely, U-shaped swipes near the nose (**AA**) showed increased angular drift, due to users’ reported aversion to the sensitive ocular area. Despite this variation, participants generally found the 4-region layout manageable, especially for midair swipes.

Variability was more pronounced for **angular L-shaped swipes** involving the nose (e.g., **AB**, **BA**, and **DA**), with users having difficulty maintaining consistent trajectories near the eye corners, particularly when swipes reached the corner of the eye. Many participants felt more comfortable executing these swipes in midair, as this approach provided a greater spatial buffer from the eyes compared to onskin swipes.

Notably, subjective ratings in the swipe-level analysis indicated low preference for certain **L-shaped swipes beyond the field of view** (i.e., between **B–C** and **C–D**), despite their low drift and high recognition accuracy. Participants expressed discomfort when onskin swipes crossed the hair region during **B–C** swipes. In midair interactions, lower ratings were primarily linked to increased wrist bending and the necessity to complete significant portions of the swipe outside the line of sight.

5.2.2 6-Region Segmentation. **6-region** layout analysis reveals finer spatial distinctions beyond the coarse patterns observed in the 4-region layout. As shown in Figure 9, swipe regions are distributed around the cheekbone focal area: **A** above the nose, **B** near the outer eye corner, **C** above the temple, **D** above the ear, **E** near the mandibular angle, and **F** near the chin. Table 3 summarizes region-level performance for swipe initiation and termination, while Tables 4a & 4b summarize swipe-shape stability across interaction spaces. Figure 10 provides representative swipe trajectories, with detailed statistical comparisons reported in Appendix B.3 & B.4.

Taken together, the region- and swipe-level analyses reveal that the ear-adjacent region **D** and the chin-adjacent regions **F** and **A** (Figure 9) produced the most stable trajectories across interaction

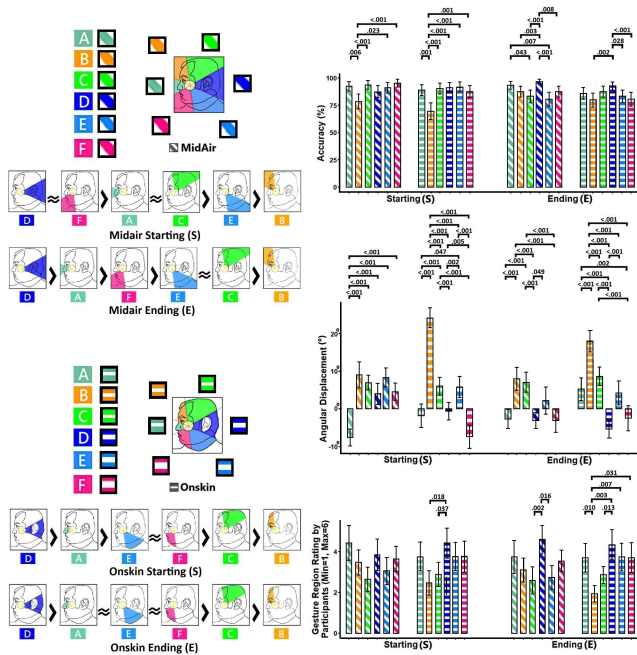


Figure 9: Gesture region layout, metrics and ranking for 6 regions.

Space	Position	Ranked Regions	Key Evidence
Midair	Start (S)	D ≈ F > A ≈ C > E > B	D / F: ↑Accuracy, ↓Displacement; stable initiation A / C: good accuracy within FOV, mild directional bias E: limited visual feedback B: ocular avoidance, lowest accuracy
	End (E)	D > A > F > C ≈ E > B	D: ↑Accuracy, ↓Displacement; stable termination A: strong visual guidance; F: moderate stability C / E: reduced separability; B: highest angular skew
Onskin	Start (S)	D > A > E ≈ F > C > B	D: ↑Ratings, ↓Displacement; most reliable tactile start A / F: combined visual + haptic feedback C: moderate stability B: ocular constraint, highest displacement
	End (E)	D > A ≈ E ≈ F > C > B	D / A: lowest displacement and highest accuracy E / F: intermediate endpoints C: reduced separability B: highest skew and lowest ratings

Table 3: Relative performance of 6 swipe regions across interaction spaces. Arrows indicate metric trends contributing to the ranking (Accuracy / ranking ↑ higher & Displacement ↓ lower is better). Accuracy = DV1, Angular Displacement = DV4, Ratings = subjective swipe region ratings.

spaces. In contrast, the eye-adjacent region B exhibited the lowest accuracy and user preference, primarily due to strong avoidance of the eye corner, particularly during onskin interactions.

Drift and avoidance patterns in regions B and C (Figure 9 and Table 3) highlight challenges in consistently distinguishing these upper-face areas midair. Their limited spatial extent and the presence of hair complicate onskin swipes, suggesting that these regions might be merged for interaction purposes.

In contrast, the lower-face regions along the cheekbone (A, D, E, and F) offer better spatial anchoring for onskin swipes. Thus, effective gesture sets should primarily focus on the lower half of the anterior-posterior axis between the nose and ear. Within this

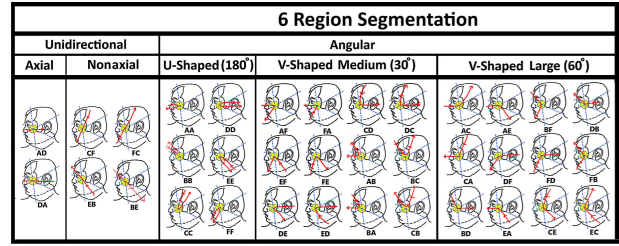


Figure 10: Visual references for individual swipe shapes in 6-region segmentation.

Swipe Group	Stable	Moderate	Unstable	Observation
Axial	AD DA	-	-	Stable on cheek-chin axis; lowest deviation
Non-axial	CF	EB	BE	Stable toward F; B, E weaker
U-shaped (180°)	AA, DD EE, FF	BB	-	U-shapes mostly stable; B / C less consistent
V-shaped Medium (30°)	CD, DC	AF, FE, EF	AB, BA	Stable near D; swipes crossing B unstable
V-shaped Large (60°)	CA, DF	AC, AE, FD	BF, DB	Stable on D-F axis; long B/C/E crossings degrade.

(a) Midair

Swipe Group	Stable	Moderate	Unstable	Observation
Axial	AD DA	-	-	Stable on cheek-chin axis; lowest deviation
Non-axial	CF	FC	EB BE	Stable toward F; swipes intersecting B unstable
U-shaped (180°)	AA, CC DD, EE	FF	BB	Stable at A/D/E/F anchors; loops at B unstable
V-shaped Medium (30°)	CD, DC, ED	FA, AF	AB, BA	Stable near D; swipes crossing B unstable
V-shaped Large (60°)	DF	AC, AE, CA	BD, BF	Stable on D-F axis; eye-crossing swipes unstable

(b) Onskin

Table 4: Swipe shape (Figure 10) stability under 6-region segmentation derived from accuracy (DV1), angular displacement (DV4), and subjective ratings.

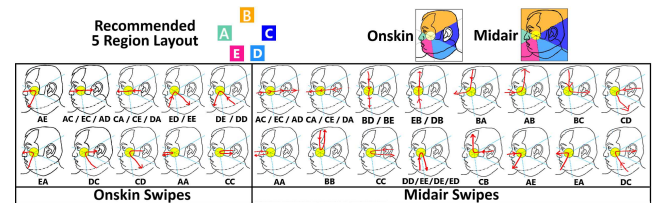


Figure 11: Proposed 5 region pattern, and a set of 26 onskin and midair swipes (10 onskin, 16 midair) after merging regions from 6-region pattern.

region, the metrics support treating E and F as distinct areas, especially for onskin interactions.

6 Discussion

6.1 Proposed 5-Region Pattern For Midair and Onskin Swipes

Drawing from consistent patterns in region preference, swipe performance, and gesture articulation across 4- and 6-region layouts, we propose a candidate 5-region pattern as a concise summary of

the above-neck, off-device earable swipe design space. Although not yet experimentally validated, this evidence-informed hypothesis reflects converging trends across region densities and interaction modalities. Addressing the density threshold observed in the 6-region layout and the recurring overlap between eye- and temple-adjacent regions, we merge the *eye* (B) and *temple* (C), while retaining the *nose* (A), *outer ear* (D), *mandible angle* (E), and *chin* (F) for both midair and onskin interaction. This layout (Figure 11) reduces region density, preserves spatial distinctiveness around key facial landmarks, and prevents over-segmentation near the eyes.

Consistent with prior work [10, 52] and our qualitative observations (Section 5.2), the mandible angle and chin remain separate due to their motor and perceptual distinctiveness. We avoid recommending onskin interaction within the merged temple region due to proximity to the eyes and potential interference from eye-wear, while noting that midair interaction remains more feasible. Finally, angular midair swipes—especially around the mandible and chin—showed consistent directional bias and low discriminability, suggesting that users did not reliably articulate or perceive some of these swipes as distinct. We therefore consolidate them into higher-level gesture abstractions aligned with dominant motion axes. Overall, the proposed 5-region pattern and set of proposed swipes (Figure 11) captures observed articulation tendencies while reducing gesture complexity.

6.2 Toward Sensing Proposed Midair and Onskin Swipes

Prior work demonstrates that both onskin and midair hand-to-face gestures can be sensed using ear-worn and in-ear platforms across diverse sensing modalities, including microphones and acoustic leakage, inertial sensing, wireless signal variation, cameras, atmospheric pressure, electric-field sensing, bio-impedance, and mmWave radar. As summarized in Section 2.1, multiple earable systems have already recognized taps, swipes, and near-face motions on or around facial regions relevant to our study. For example, *EarBuddy* [90], *EarAcE* [10], and *OESense* [52] recognize onskin interactions on regions such as the cheekbone, temple, mandible angle, and chin, while *FaceSense* [41] explores camera placement near the earlobe to capture hand-to-face interaction. Similarly, midair sensing around the face has been demonstrated using BLE signal strength (*EarGest* [2]), acoustic leakage (*LeakyFeeder* [92], *EarHover* [74], *MAF* [93]), IMUs (*UiEar* [97]), atmospheric pressure [36], and in-ear mmWave radar (*TinyssimoRadar* [68]). Recent and forthcoming industry efforts further suggest the practical viability of near-face midair sensing on consumer earables [35, 38].

Our candidate 5-region layout (Section 6.1) was derived from observed articulation and region-level tendencies, emphasizing facial areas that already have precedent in the sensing literature (e.g., nose / cheek-adjacent regions, temple, tragus / outer ear, mandible angle, and chin). Because the 5-region configuration itself was not evaluated as an experimental condition, we treat it as an evidence-informed design hypothesis rather than a validated sensing layout; consequently, the sensing implications of heterogeneous region sizes and boundaries remain to be empirically examined.

Nevertheless, the proposed gesture vocabulary is compatible with recognition strategies used in prior systems. For onskin interaction, many proposed unidirectional and angular swipes can be modeled as trajectories across a small set of touch-relevant landmarks. For example, a long onskin swipe (e.g., EA in Figure 11) can be interpreted as a sequence of contacts and short transitions (chin → cheekbone → nose), where each subregion has independently demonstrated discriminative sensing characteristics in prior work. Similarly, for midair interaction, several systems already recognize directional swipes and near-face motion primitives; angular swipes may therefore be decomposed into short axial segments (e.g., a vertical segment followed by a horizontal segment), enabling recognition pipelines that compose primitives into higher-level gestures.

Taken together, these findings and prior sensing systems suggest a plausible path toward recognizing the 26 proposed midair and onskin swipes discussed in Section 6.1, while also motivating follow-up sensing studies to validate region layouts, shape-level constraints, and end-to-end robustness on practical earable hardware. Consistent with prior work on imaginary and off-device interfaces [17], our contributions ground candidate gesture designs in empirically observed articulation patterns and sensing capabilities already demonstrated in the earable interaction literature.

7 Limitation and Future Work

While this work synthesizes a candidate 5-region layout and a corresponding set of 26 unidirectional and angular midair and onskin swipes for off-device earable interaction, the configuration itself was not evaluated as an experimental condition and therefore remains an evidence-informed design hypothesis. Additionally, our study was conducted under seated, controlled conditions to isolate spatial effects around the face, limiting generalizability to mobile everyday contexts where motor variability, comfort, and social acceptability may differ.

As follow-up work, we plan to empirically examine how region size, shape, and heterogeneity affect performance, learnability, recognition accuracy, and robustness across static and mobile settings. We further plan to validate key swipe-shape design decisions (e.g., consolidating angular midair swipes and separating midair and onskin interaction) through targeted swipe-level evaluations informed by prior findings on location-specific tap, swipe, and pinch gestures in above-neck earable interaction [6]. Additionally, while emerging in-ear sensing platforms show promise [68, 69], we plan to examine recognition accuracy, latency, and robustness to articulation variability for the proposed gestures under practical sensing conditions.

Beyond validating the proposed interaction space, we also aim to explore hybrid hand-to-face techniques, similar to *Air+Touch* [12], that combine stable onskin anchors (e.g., the cheekbone focal region) with midair motion for off-device earable interaction. From an application perspective, we also plan to investigate how off-device earable swipe gestures integrate with XR systems, remote display interaction, and earable voice interfaces such as *Siri* on *Apple AirPods* [11], where hand-to-face gestures may complement speech and lightweight manual input to support multimodal interaction without sustained visual attention [89].

8 Conclusion

This work investigates unimanual hand-to-face unidirectional and angular swipe gestures as an off-device input modality for in-ear earables. Through controlled analysis across interaction spaces, region densities, and region locations, we characterize how users naturally articulate these single-stroke swipes, and how these behaviors relate to performance and region preference for swipe gestures. Based on these empirical observations, we synthesize a candidate 5-region pattern and a reduced set of midair and onskin unidirectional and angular swipes to structure and reason about the design space of off-device manual earable interaction. This synthesis represents an evidence-informed design hypothesis rather than a validated gesture configuration. Overall, this work advances understanding of hand-to-face swipe behavior for off-device interaction with in-ear earable device like *Apple Airpod*, and provides a grounded foundation for future research on validating region layouts, refining shape-aware gesture mappings, and exploring sensing and application-level implications. Although grounded in earable interaction, the contributions of this work generalize to a wider class of above-neck, contact-free gestural interaction techniques for head-worn systems. By characterizing how users perform and structure hand-to-face swipes, this work provides actionable insight for the design of off-device gestural input across earables, head-mounted displays, and other head-worn wearables.

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*). ACM, New York, NY, USA, 193–202. doi:10.1145/2628363.2628381
- [2] Khaled Alkielek, Khaled A. Harras, and Moustafa Youssef. 2022. EarGest: Hand Gesture Recognition with Earables. In *Proceedings of the 2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking* (Stockholm, Sweden) (*SECON '22*). IEEE, Piscataway, NJ, USA, 91–99. doi:10.1109/SECON55815.2022.9918622
- [3] Khaled Alkielek, Moustafa Youssef, and Khaled A. Harras. 2023. EarBender: Enabling Rich IMU-based Natural Hand-to-Ear Interaction in Commodity Earables. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing* (Cancun, Quintana Roo, Mexico) (*UbiComp/ISWC '23 Adjunct*). ACM, New York, NY, USA, 333–338. doi:10.1145/3594739.3610671
- [4] Lisa Anthony and Jacob O. Wobbrock. 2010. A Lightweight Multistroke Recognizer for User Interface Prototypes. In *Proceedings of Graphics Interface 2010* (Ottawa, ON, Canada) (*GI '10*). CIPS, Toronto, ON, Canada, 245–252. doi:10.5555/1839214.1839258
- [5] Apple Inc. 2025. *Introducing AirPods Pro 3, the Ultimate Audio Experience*. Apple Inc. Retrieved May 18, 2026 from <https://www.apple.com/newsroom/2025/09/introducing-airpods-pro-3-the-ultimate-audio-experience/>
- [6] 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. doi:10.1145/3643513
- [7] Felix Barroso, Norbert Freedman, and Stanley Grand. 1980. Self-Touching, Performance, and Attentional Processes. *Percept. Mot. Skills* 50, 3_suppl (June 1980), 1083–1089. doi:10.2466/pms.1980.50.3c.1083
- [8] Abdelkareem Bedri, David Byrd, Peter Presti, Himanshu Sahni, Zehua Gue, and Thad Starner. 2015. Stick It in Your Ear: Building an In-Ear Jaw Movement Sensor. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (Osaka, Japan) (*UbiComp/ISWC'15 Adjunct*). ACM, New York, NY, USA, 1333–1338. doi:10.1145/2800835.2807933
- [9] Xiang Cao and Shumin Zhai. 2007. Modeling Human Performance of Pen Stroke Gestures. In *Proceedings of the 2007 SIGCHI Conference on Human Factors in Computing Systems* (San Jose, CA, USA) (*CHI '07*). ACM, New York, NY, USA, 1495–1504. doi:10.1145/1240624.1240850
- [10] 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 (June 2023), 23 pages. doi:10.1145/3596242
- [11] Tao Chen, Yongjie Yang, Chonghao Qiu, Xiaoran Fan, Xiuzhen Guo, and Longfei Shangquan. 2024. Enabling Hands-Free Voice Assistant Activation on Earphones. In *Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services* (Minato-ku, Tokyo, Japan) (*MOBISYS '24*). ACM, New York, NY, USA, 155–168. doi:10.1145/3643832.3661890
- [12] Xiang 'Anthony' Chen, Julia Schwarz, Chris Harrison, Jennifer Mankoff, and Scott E. Hudson. 2014. Air+touch: Interweaving Touch & In-air Gestures. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, HI, USA) (*UIST '14*). ACM, New York, NY, USA, 519–525. doi:10.1145/2642918.2647392
- [13] 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. doi:10.1145/3427314
- [14] Wen-Wei Cheng and Liwei Chan. 2024. BodyTouch: Investigating Eye-Free, On-Body and Near-Body Touch Interactions with HMDs. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4, Article 152 (Jan. 2024), 22 pages. doi:10.1145/3631426
- [15] 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 (July 2022), 32 pages. doi:10.1145/3534597
- [16] 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 Event, UK) (*HotMobile '21*). ACM, New York, NY, USA, 147–153. doi:10.1145/3446382.3450216
- [17] Niloofer 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*). ACM, New York, NY, USA, 27–34. doi:10.1145/2325616.2325623
- [18] Pouriya Amini Digehsara, Joao Vitor Possamai de Menezes, Christoph Wagner, Michael Bärhold, Petr Schaffer, Dirk Plettemeier, and Peter Birkholz. 2022. A User-Friendly Headset for Radar-Based Silent Speech Recognition. *Proc. Interspeech* 7, 21 (Sept. 2022), 4835–4839. doi:10.21437/Interspeech.2022-10090
- [19] David Döbelstein, Gabriel Haas, and Enrico Rukzio. 2017. The Effects of Mobility, Encumbrance, and (Non-)dominant Hand on Interaction with Smartwatches. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, HI, USA) (*ISWC '17*). ACM, New York, NY, USA, 90–93. doi:10.1145/3123021.3123033
- [20] John D Eastwood, Daniel Smilek, and Philip M Merikle. 2003. Negative Facial Expression Captures Attention and Disrupts Performance. *Percept. Psychophys.* 65, 3 (April 2003), 352–358. doi:10.3758/BF03194566
- [21] Paul Ekman and Wallace V. Friesen. 1972. Hand Movements. *J. Commun.* 22, 4 (Dec. 1972), 353–374. doi:10.1111/j.1460-2466.1972.tb00163.x
- [22] Shariff AM Faleel, Michael Gammon, Kevin Fan, Da-Yuan Huang, Wei Li, and Pourang Irani. 2021. HPUI: Hand Proximate User Interfaces for One-Handed Interactions on Head Mounted Displays. *IEEE Trans. Vis. Comput. Graph.* 27, 11 (Aug. 2021), 4215–4225. doi:10.1109/TVCG.2021.3106493
- [23] Jonny Farrington, Vanessa Oni, Chi Ming Kan, and Leo Poll. 1999. Co-Modal Browser-An Interface for Wearable Computers. In *Digest of Papers. Third International Symposium on Wearable Computers* (San Francisco, CA, USA) (*ISWC '99*). IEEE, Piscataway, NJ, USA, 45–51. doi:10.1109/ISWC.1999.806644
- [24] 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, UK) (*EarComp '19*). ACM, New York, NY, USA, 8–13. doi:10.1145/3345615.3361131
- [25] Shkurta Gashi, Aaqib Saeed, Alessandra Vicini, Elena Di Lascio, and Silvia Santini. 2021. Hierarchical Classification and Transfer Learning to Recognize Head Gestures and Facial Expressions Using Earbuds. In *Proceedings of the 2021 International Conference on Multimodal Interaction* (Montreal, QC, Canada) (*ICMI '21*). ACM, New York, NY, USA, 168–176. doi:10.1145/3462244.3479921
- [26] 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. doi:10.1145/3596474
- [27] Tovi Grossman, Xiang Anthony Chen, and George Fitzmaurice. 2015. Typing on Glasses: Adapting Text Entry to Smart Eyewear. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) (*MobileHCI '15*). ACM, New York, NY, USA, 144–152. doi:10.1145/2785830.2785867
- [28] 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, NY, USA) (*UIST '10*). ACM, New York, NY, USA, 3–12. doi:10.1145/1866029.1866033

- [29] Arman Hafizi, Jay Henderson, Ali Neshati, Wei Zhou, Edward Lank, and Daniel Vogel. 2023. In-vehicle Performance and Distraction for Midair and Touch Directional Gestures. In *Proceedings of the 2023 SIGCHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). ACM, New York, NY, USA, Article 316, 13 pages. doi:10.1145/3544548.3581335
- [30] Jinni A Harrigan, John R Kues, John J Steffen, and Robert Rosenthal. 1987. Self-Touching and Impressions of Others. *Pers. Soc. Psychol. Bull.* 13, 4 (Dec. 1987), 497–512. doi:10.1177/0146167287134007
- [31] Hayati Havlucu, Mehmet Yarkin Ergin, İdil Bostan, Oğuz Turan Buruk, Tilbe Gökşun, and Oğuzhan Özcan. 2017. It Made More Sense: Comparison of User-Elicited On-skin Touch and Freehand Gesture Sets. In *Proceedings of the 5th International Conference on Distributed, Ambient and Pervasive Interactions, held as part of HCI International 2017* (Vancouver, BC, Canada) (DAPI '17). Springer, Cham, Switzerland, 159–171. doi:10.1007/978-3-319-58697-7_11
- [32] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasi, and Pourang Irani. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-Air Interactions. In *Proceedings of the 2014 SIGCHI Conference on Human Factors in Computing Systems* (Toronto, ON, Canada) (CHI '14). ACM, New York, NY, USA, 1063–1072. doi:10.1145/2556288.2557130
- [33] 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). ACM, New York, NY, USA, 12–23. doi:10.1145/3490632.3490636
- [34] Jingyang Hu, Hongbo Jiang, Zhu Xiao, Siyu Chen, Schahram Dustdar, and Jiangchuan Liu. 2024. HeadTrack: Real-Time Human-Computer Interaction via Wireless Earphones. *IEEE J. Sel. Areas Commun.* 42, 4 (April 2024), 990–1002. doi:10.1109/JSAC.2023.3345381
- [35] Roland Hutchinson. 2026. Beyond Audio: Why the AirPods Ultra's New Infrared Sensors Are a Game-Changer. Retrieved May 19, 2026 from <https://www.geeky-gadgets.com/apple-airpods-ultra-cameras-ai/>
- [36] Koki Iguma, Kazuya Muroa, and Hiroki Watanabe. 2023. Input Interface with Touch and Non-touch Interactions using Atmospheric Pressure for Hearable Devices. In *Proceedings of the 2023 ACM International Symposium on Wearable Computers* (Cancun, Quintana Roo, Mexico) (ISWC '23). ACM, New York, NY, USA, 1–5. doi:10.1145/3594738.3611354
- [37] Poika Isokoski. 2001. Model for Unistroke Writing Time. In *Proceedings of the 2001 SIGCHI Conference on Human Factors in Computing Systems* (Seattle, WA, USA) (CHI '01). ACM, New York, NY, USA, 357–364. doi:10.1145/365024.365299
- [38] Benjamin G. Jackson, Brenton A. Baugh, David H. Bloom, Gemma A. Roper, Karlin Y. Bark, and Thomas S. Hulbert. 2023. Wearable Interactive Audio Device. <https://patents.google.com/patent/US11743623B2> Patent No. US11743623B2, Filed July 13th., 2020, Issued Aug. 29th., 2023.
- [39] 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 (July 2022), 28 pages. doi:10.1145/3534613
- [40] Yincheng Jin, Yang Gao, Yanjun Zhu, Wei Wang, Jiyang Li, Seokmin Choi, Zhangyu Li, Jagmohan Chauhan, Anind K. Dey, and Zhanpeng Jin. 2021. Soni-cASL: An Acoustic-based Sign Language Gesture Recognizer Using Earphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 67 (June 2021), 30 pages. doi:10.1145/3463519
- [41] 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 (Sept. 2021), 27 pages. doi:10.1145/3478129
- [42] Prerna Khanna, Tanmay Srivastava, Shijia Pan, Shubham Jain, and Phuc Nguyen. 2021. JawSense: Recognizing Unvoiced Sound using a Low-cost Ear-worn System. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications* (Virtual Event, UK) (HotMobile '21). ACM, New York, NY, USA, 44–49. doi:10.1145/3446382.3448363
- [43] 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). ACM, New York, NY, USA, Article 27, 6 pages. doi:10.1145/3098279.3098538
- [44] 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 SIGCHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). ACM, New York, NY, USA, 1–19. doi:10.1145/3313831.3376162
- [45] Barry Kolle, Sven Kratz, and Anthony Dunnigan. 2014. Exploring Gestural Interaction in Smart Spaces Using Head-Mounted Devices with Ego-centric Sensing. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction* (Honolulu, HI, USA) (SUI '14). ACM, New York, NY, USA, 40–49. doi:10.1145/2659766.2659781
- [46] Yen Lee Angela Kwok, Jan Gralton, and Mary-Louise McLaws. 2015. Face Touching: A Frequent Habit That Has Implications for Hand Hygiene. *Am. J. Infect. Control* 43, 2 (Feb. 2015), 112–114. doi:10.1016/j.ajic.2014.10.015
- [47] Jae Yeol Lee, Gue Won Rhee, and Dong Woo Seo. 2010. Hand Gesture-based Tangible Interactions for Manipulating Virtual Objects in a Mixed Reality Environment. *Int. J. Adv. Manuf. Technol.* 51 (April 2010), 1069–1082. doi:10.1007/s00170-010-2671-x
- [48] Luis A. Leiva, Sunjun Kim, Wenzhe Cui, Xiaojun Bi, and Antti Oulasvirta. 2021. How We Swipe: A Large-scale Shape-writing Dataset and Empirical Findings. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) (MobileHCI '21). ACM, New York, NY, USA, Article 11, 13 pages. doi:10.1145/3447526.3472059
- [49] Ning Li, Tuoyang Zhou, Yingwei Zhou, Chen Guo, Deqiang Fu, Xiaoqing Li, and Zijing Guo. 2019. Research on Human-Computer Interaction Mode of Speech Recognition Based on Environment Elements of Command and Control System. In *Proceedings of the 2019 5th International Conference on Big Data and Information Analytics* (Kunming, China) (BigDIA '19). IEEE, Piscataway, NJ, USA, 170–175. doi:10.1109/BigDIA.2019.8802812
- [50] Wenmin Li, Xueyi Wan, Yanwei Shi, Nailang Yao, Ci Wang, and Zaifeng Gao. 2021. Depth and Breadth of Pie Menus for Mid-air Gesture Interaction. *Int. J. Hum.-Comput. Interact.* 37, 2 (Aug. 2021), 131–140. doi:10.1080/10447318.2020.1809245
- [51] Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, Suranga Nanayakkara, and Max Mühlhäuser. 2014. EarPut: Augmenting Ear-Worn Devices for Ear-Based Interaction. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (Sydney, New South Wales, Australia) (OzCHI '14). ACM, New York, NY, USA, 300–307. doi:10.1145/2686612.2686655
- [52] Dong Ma, Andrea Ferlini, and Cecilia Mascolo. 2021. OESense: Employing Occlusion Effect for in-Ear Human Sensing. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services* (Virtual Event, WI, USA) (MobiSys '21). ACM, New York, NY, USA, 175–187. doi:10.1145/3458864.3467680
- [53] C. Metzger, M. Anderson, and T. Starner. 2004. FreeDigitizer: A Contact-Free Device for Gesture Control. In *Proceedings of the 2004 8th IEEE International Symposium on Wearable Computers* (Arlington, VA, USA) (ISWC '04, Vol. 1). IEEE, Piscataway, NJ, USA, 18–21. doi:10.1109/ISWC.2004.23
- [54] Microsoft. 2024. Azure Kinect DK. <https://azure.microsoft.com/en-us/products/kinect-dk>
- [55] Khairul Khaizi Mohd Shariff, Auni Nadiah Yusni, Mohd Adli Md Ali, Megat Syahirul Amin Megat Ali, Megat Zuhairi Megat Tajuddin, and M A A Youmis. 2022. CW Radar Based Silent Speech Interface Using CNN. In *Proceedings of the 2022 IEEE Symposium on Wireless Technology & Applications* (Kuala Lumpur, Malaysia) (ISWTA '22). IEEE, Piscataway, NJ, USA, 76–81. doi:10.1109/ISWTA55313.2022.9942730
- [56] 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, Article e0213677 (March 2019), 19 pages. doi:10.1371/journal.pone.0213677
- [57] Mark Nicas and Daniel Best. 2008. A Study Quantifying the Hand-to-Face Contact Rate and Its Potential Application to Predicting Respiratory Tract Infection. *J. Gen. Psychol.* 5, 6 (April 2008), 347–352. doi:10.1080/15459620802003896
- [58] AUKEY Official. 2025. AUKEY EP-T27 Soundstream Wireless Earbuds Noise Cancelling IPX7 Waterproof Black. <https://www.aukey.com/products/aukey-soundstream-wireless-earbuds-noise-cancelling-ipx7-waterproof-black>
- [59] NaturalPoint, Inc. DBA OptiTrack. 2024. Optitrack: The Best Just Got Better. <https://www.optitrack.com/about/>
- [60] Laxmi Pandey, Khalad Hasan, and Ahmed Sabbir Arif. 2021. Acceptability of Speech and Silent Speech Input Methods in Private and Public. In *Proceedings of the 2021 SIGCHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). ACM, New York, NY, USA, Article 251, 13 pages. doi:10.1145/3411764.3445430
- [61] Francisco Perella-Holfeld, Shariff AM Faleel, and Pourang Irani. 2023. Evaluating Design Guidelines for Hand-Proximate User Interfaces. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) (DIS '23). ACM, New York, NY, USA, 1159–1173. doi:10.1145/3563657.3596117
- [62] Sen Qiu, Zhelong Wang, Hongyu Zhao, and Huosheng Hu. 2016. Using Distributed Wearable Sensors to Measure and Evaluate Human Lower Limb Motions. *IEEE Trans. Instrum. Meas.* 65, 4 (April 2016), 939–950. doi:10.1109/TIM.2015.2504078
- [63] Meera Radhakrishnan, Kushaan Misra, and V. Ravichandran. 2021. Applying "Earable" Inertial Sensing for Real-time Head Posture Detection. In *Proceedings of the 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events* (Virtual Event, Germany) (PerCom Workshops '21). IEEE, Piscataway, NJ, USA, 176–181. doi:10.1109/PerComWorkshops51409.2021.9430988
- [64] 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. doi:10.1145/3567710

- [65] Vera Remizova, Yulia Gizatdinova, and Veikko Surakka. 2022. Midair Gestural Techniques for Translation Tasks in Large-Display Interaction. *Adv. Hum.-Comput. Interact.* 2022, 1, Article 9362916 (July 2022), 13 pages. doi:10.1155/2022/9362916
- [66] 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 (Sept. 2022), 28 pages. doi:10.1145/3546736
- [67] Tobias Röddiger, Christopher Clarke, Paula Breiitling, 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 (Sept. 2022), 1–57. doi:10.1145/3550314
- [68] Andrea Ronco, Philipp Schilk, and Michele Magno. 2024. TinyssimoRadar: In-Ear Hand Gesture Recognition with Ultra-Low Power mmWave Radars. In *Proceedings of the 2024 IEEE/ACM Ninth International Conference on Internet-of-Things Design and Implementation* (Hong Kong, Hong Kong SAR) (IoTDI '24). IEEE, Piscataway, NJ, USA, 192–202. doi:10.1109/IoTDI61053.2024.00021
- [69] Yukina Sato, Takashi Amesaka, Takumi Yamamoto, Hiroki Watanabe, and Yuta Sugiura. 2024. Exploring User-Defined Gestures as Input for Hearables and Recognizing Ear-Level Gestures with IMUs. *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 258 (Sept. 2024), 23 pages. doi:10.1145/3676503
- [70] Marcos Serrano, Barrett M. Ens, and Pourang P. Irani. 2014. Exploring the Use of Hand-to-face Input for Interacting with Head-worn Displays. In *Proceedings of the 2014 SIGCHI Conference on Human Factors in Computing Systems* (Toronto, ON, Canada) (CHI '14). ACM, New York, NY, USA, 3181–3190. doi:10.1145/2556288.2556984
- [71] Kingzhe Song, Kai Huang, and Wei Gao. 2022. FaceListener: Recognizing Human Facial Expressions via Acoustic Sensing on Commodity Headphones. In *Proceedings of the 2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks* (Milan, Italy) (IPSN '22). IEEE, Piscataway, NJ, USA, 145–157. doi:10.1109/IPSN54338.2022.00019
- [72] Tanmay Srivastava, Purna Khanna, Shijia Pan, Phuc Nguyen, and Shubham Jain. 2022. Mutelt: Jaw Motion Based Unvoiced Command Recognition Using Earable. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 140 (Sept. 2022), 26 pages. doi:10.1145/3550281
- [73] Wei Sun, Franklin Mingzhe Li, Benjamin Steeper, Songlin Xu, Feng Tian, and Cheng Zhang. 2021. TeethTap: Recognizing Discrete Teeth Gestures Using Motion and Acoustic Sensing on an Earpiece. In *Proceedings of the 26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) (IUI '21). ACM, New York, NY, USA, 161–169. doi:10.1145/3397481.3450645
- [74] Shunta Suzuki, Takashi Amesaka, Hiroki Watanabe, Buntarou Shizuki, and Yuta Sugiura. 2024. EarHover: Mid-Air Gesture Recognition for Hearables Using Sound Leakage Signals. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) (UIST '24). ACM, New York, NY, USA, Article 129, 13 pages. doi:10.1145/3654777.3676367
- [75] Huawei Tu, Weiyang Huan, Xingdong Yang, Xiangshi Ren, and Feng Tian. 2021. ArmMenu: Command Input on Distant Displays with Proprioception-Based Lateral Arm Movements. *Behav. Inf. Technol.* 40, 13 (2021), 1428–1447. doi:10.1080/0144929X.2020.1757760
- [76] Vicon Motion Systems Ltd UK. 2023. Vicon: What is Motion Capture - Intelligence in Motion. <https://www.vicon.com/about-us/what-is-motion-capture/>
- [77] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. 2012. Gestures as Point Clouds: A SP Recognizer for User Interface Prototypes. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction* (Santa Monica, CA, USA) (ICMI '12). ACM, New York, NY, USA, 273–280. doi:10.1145/2388676.2388732
- [78] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. 2013. Relative Accuracy Measures for Stroke Gestures. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (Sydney, Australia) (ICMI '13). ACM, New York, NY, USA, 279–286. doi:10.1145/2522848.2522875
- [79] Radu-Daniel Vatavu and Ovidiu-Ciprian Ungurean. 2022. Understanding Gesture Input Articulation with Upper-Body Wearables for Users with Upper-Body Motor Impairments. In *Proceedings of the 2022 SIGCHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). ACM, New York, NY, USA, Article 2, 16 pages. doi:10.1145/3491102.3501964
- [80] Radu-Daniel Vatavu, Daniel Vogel, Gery Casiez, and Laurent Grisoni. 2011. Estimating the Perceived Difficulty of Pen Gestures. In *Proceedings of the 13th IFIP TC 13 International Conference, Part 1* (Lisbon, Portugal) (Interact '11). Springer, Berlin-Heidelberg, Germany, 89–106. doi:10.1007/978-3-642-23771-3_9
- [81] M. Vella. 2008. *Anatomy for Strength and Fitness Training* (1st ed.). New Holland, London, UK. https://books.google.ca/books?id=Q3_MvAEACAAJ
- [82] Dhruv Verma, Sejal Bhalha, Dhruv Sahnjan, Jainendra Shukla, and Aman Parnami. 2021. ExpressEar: Sensing Fine-Grained Facial Expressions with Earables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 129 (Sept. 2021), 28 pages. doi:10.1145/3478085
- [83] Lichen Wang, Bin Sun, Joseph Robinson, Taotao Jing, and Yun Fu. 2020. EV-Action: Electromyography-Vision Multi-Modal Action Dataset. In *Proceedings of the 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition* (Buenos Aires, Argentina) (FG '20). IEEE, Piscataway, NJ, USA, 160–167. doi:10.1109/FG47880.2020.00018
- [84] Yueting Weng, Chun Yu, Yingtian Shi, Yuhang Zhao, Yukang Yan, and Yuanchun Shi. 2021. FaceSight: Enabling Hand-to-Face Gesture Interaction on AR Glasses with a Downward-Facing Camera Vision. In *Proceedings of the 2021 SIGCHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). ACM, New York, NY, USA, Article 10, 14 pages. doi:10.1145/3411764.3445484
- [85] Markus L. Wittorf and Mikkel R. Jakobsen. 2016. Eliciting Mid-Air Gestures for Wall-Display Interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (Gothenburg, Sweden) (NordCHI '16). ACM, New York, NY, USA, Article 3, 4 pages. doi:10.1145/2971485.2971503
- [86] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-Defined Gestures for Surface Computing. In *Proceedings of the 2009 SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) (CHI '09). ACM, New York, NY, USA, 1083–1092. doi:10.1145/1518701.1518866
- [87] Jacob O. Wobbrock, Andrew D. Wilson, and Yang Li. 2007. Gestures without libraries, toolkits or training: a \$1 recognizer for user interface prototypes. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology* (Newport, RI, USA) (UIST '07). ACM, New York, NY, USA, 159–168. doi:10.1145/1294211.1294238
- [88] 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 SIGCHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). ACM, New York, NY, USA, 1–11. doi:10.1145/3313831.3376393
- [89] Liwei Wu, Ben Lafreniere, Tovi Grossman, Thomas White, and Stephanie Santosa. 2024. Body Language for VUIs: Exploring Gestures to Enhance Interactions with Voice User Interfaces. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (DIS '24). ACM, New York, NY, USA, 133–150. doi:10.1145/3643834.3660691
- [90] 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 SIGCHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). ACM, New York, NY, USA, 1–14. doi:10.1145/3313831.3376836
- [91] Shota Yamanaka, Hiroki Usuba, and Junichi Sato. 2024. Behavioral Differences Between Tap and Swipe: Observations on Time, Error, Touch-point Distribution, and Trajectory for Tap-and-swipe Enabled Targets. In *Proceedings of the 2024 SIGCHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). ACM, New York, NY, USA, Article 549, 12 pages. doi:10.1145/3613904.3642272
- [92] Yongjie Yang, Tao Chen, Zhenlin An, Shirui Cao, Xiaoran Fan, and Longfei Shangguan. 2025. LeakyFeeder: In-Air Gesture Control Through Leaky Acoustic Waves. In *Proceedings of the 23rd ACM Conference on Embedded Networked Sensor Systems* (Irvine, CA, USA) (SenSys '25). ACM, New York, NY, USA, 144–157. doi:10.1145/3715014.3722054
- [93] Yongjie Yang, Tao Chen, Yujing Huang, Xiuzhen Guo, and Longfei Shangguan. 2024. MAF: Exploring Mobile Acoustic Field for Hand-to-Face Gesture Interactions. In *Proceedings of the 2024 SIGCHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). ACM, New York, NY, USA, Article 638, 20 pages. doi:10.1145/3613904.3642437
- [94] 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 (March 2023), 28 pages. doi:10.1145/3580838
- [95] Shijia Zhang, Taiting Lu, Hao Zhou, Yilin Liu, Runze Liu, and Mahanth Gowda. 2023. I Am an Earphone and I Can Hear My User's Face: Facial Landmark Tracking Using Smart Earphones. *ACM Trans. Internet Things* 5, 1, Article 1 (Dec. 2023), 29 pages. doi:10.1145/3614438
- [96] 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, MA, USA) (SenSys '22). ACM, New York, NY, USA, 847–848. doi:10.1145/3560905.3568075
- [97] Guangrong Zhao, Yiran Shen, Feng Li, Lei Liu, Lizhen Cui, and Hongkai Wen. 2025. Ui-Ear: On-Face Gesture Recognition Through On-Ear Vibration Sensing. *IEEE Trans. Mob. Comput.* 24, 3 (2025), 1482–1495. doi:10.1109/TMC.2024.3480216
- [98] Quan Zhou, Bin Fang, Jianhua Shan, Fuchun Sun, and Di Guo. 2020. A Survey of the Development of Wearable Devices. In *Proceedings of the 2020 5th International Conference on Advanced Robotics and Mechatronics* (Shenzhen, China) (ICARM '20). IEEE, Piscataway, NJ, USA, 198–203. doi:10.1109/ICARM49381.2020.9195351
- [99] Peizhao Zhu, Yongpan Zou, Wenyuan Li, and Kaishun Wu. 2023. CHAR: Composite Head-body Activities Recognition with A Single Earable Device. In *Proceedings of the 2023 IEEE International Conference on Pervasive Computing and Communications* (Atlanta, GA, USA) (PerCom '23). IEEE, Piscataway, NJ, USA, 212–221. doi:10.1109/PERCOM56429.2023.10099218

Appendices

A Detailed Statistical Results for Section 5.1

A.1 RQ1

Metric	Regions	Shapiro-Wilk Normality Test				Wilcoxon Signed Rank Test (Paired Data)				Paired t-test			
		Midair		Onskin		W	Z	p	r	t(23)	p	Cohen's d	
		W	p	W	p								
Accuracy (DV1)	4	.258	<.001	.144	<.001	148.500	-2.475	.014	.089	-	-	-	
	6	.348	<.001	.404	<.001	5922.000	2.366	.018	.056	-	-	-	
	8	.423	<.001	.439	<.001	37233.000	.875	.381	.015	-	-	-	
Time (DV2)	4	.865	<.001	.924	<.001	13924.000	-10.584	<.001	.382	-	-	-	
	6	.925	<.001	.951	<.001	79484.000	-14.631	<.001	.352	-	-	-	
	8	.944	<.001	.970	<.001	218413.000	-21.360	<.001	.385	-	-	-	
Path Length (DV3)	4	.945	<.001	.891	<.001	61402.000	11.230	<.001	.405	-	-	-	
	6	.961	<.001	.934	<.001	280603.000	12.778	<.001	.307	-	-	-	
	8	.968	<.001	.893	<.001	919425.000	18.935	<.001	.342	-	-	-	
Angular Displacement (DV4)	S	4	.946	<.001	.956	<.001	27268.000	-2.446	.016	.088	-	-	-
		6	.979	<.001	.977	<.001	148493.000	.320	.841	.007	-	-	-
		8	.952	<.001	.895	<.001	439705.000	4.543	<.001	.082	-	-	-
	E	4	.961	<.001	.966	<.001	30875.000	-.247	.849	.009	-	-	-
		6	.976	<.001	.975	<.001	120336.000	-4.068	<.001	.097	-	-	-
		8	.980	<.001	.902	<.001	411172.000	-7.317	<.001	.132	-	-	-
Mental Demand	4	.815	<.001	.901	.023	62.500	-1.585	.195	.229	-	-	-	
	6	.914	.043	.924	.073	105.500	-.100	1.000	.014	-	-	-	
	8	.929	.006	.905	.028	117.000	-.672	.532	.097	-	-	-	
Physical Demand	4	.801	<.001	.848	<.001	79.500	.392	.905	.056	-	-	-	
	6	.914	.043	.950	.276	128.500	1.019	.389	.147	-	-	-	
	8	.934	.117	.955	.348	-	-	-	.671	.508	.137		
Temporal Demand	4	.921	.060	.899	.020	131.500	1.106	.330	.160	-	-	-	
	6	.958	.408	.958	.240	-	-	-	.650	.522	.133		
	8	.955	.343	.964	.515	-	-	-	.487	.631	.010		
Performance	4	.874	.006	.870	.005	154.000	1.018	.379	.147	-	-	-	
	6	.900	.021	.918	.052	144.000	1.349	.150	.195	-	-	-	
	8	.914	.042	.969	.648	156.500	1.389	.159	.200	-	-	-	
Effort	4	.918	.052	.887	.012	152.500	1.285	.199	.225	-	-	-	
	6	.914	.044	.920	.058	131.000	1.093	.338	.158	-	-	-	
	8	.909	.034	.926	.027	144.000	1.234	.150	.178	-	-	-	
Frustration	4	.799	<.001	.806	<.001	63.000	.221	.886	.031	-	-	-	
	6	.817	<.001	.920	.059	86.500	.460	.746	.066	-	-	-	
	8	.931	.102	.913	.040	89.500	.682	.553	.098	-	-	-	

Table A.1: Statistical comparison of swipe performance metrics across midair and onskin spaces for different region segmentation patterns (RQ1). Figure 6 highlights statistically significant pairwise differences.

A.2 RQ2

Metric	Interaction Space	Shapiro-Wilk Normality Test						Friedman Test (Matched Group)		Post-Hoc Wilcoxon Test with Bonferroni Correction For Pairwise Comparison			One-way Repeated Measures Anova for Pairwise Comparison			Post-Hoc Pairwise t-test with Bonferroni Correction		
		4 Regions		6 Regions		8 Regions		χ^2	p	p_{4-6}	p_{6-8}	p_{4-8}	$F(2,46)$	p	η^2	p_{4-6}	p_{6-8}	p_{4-8}
		W	p	W	p	W	p											
Accuracy (DV1)	Midair	.258	<.001	.348	<.001	.423	<.001	57.901	<.001	.064	.006	<.001	-	-	-	-	-	-
	Onskin	.144	<.001	.404	<.001	.439	<.001	164.760	<.001	<.001	.370	<.001	-	-	-	-	-	
Time (DV2)	Midair	.865	<.001	.925	<.001	.944	<.001	55.202	<.001	.034	.022	<.001	-	-	-	-	-	
	Onskin	.924	<.001	.951	<.001	.970	<.001	20.333	<.001	1.000	<.001	.026	-	-	-	-	-	
Path Length (DV3)	Midair	.945	<.001	.961	<.001	.968	<.001	39.470	<.001	.080	.002	1.000	-	-	-	-	-	
	Onskin	.891	<.001	.934	<.001	.893	<.001	16.353	<.001	.238	.161	.001	-	-	-	-	-	
Angular Displacement (DV4)	S	Midair	.946	<.001	.979	<.001	.952	<.001	148.490	<.001	<.001	<.001	<.001	-	-	-	-	-
		Onskin	.956	<.001	.977	<.001	.895	<.001	148.240	<.001	<.001	<.001	<.001	-	-	-	-	
		Midair	.961	<.001	.976	<.001	.980	<.001	118.210	<.001	<.001	<.001	<.001	-	-	-	-	
	E	Onskin	.966	<.001	.975	<.001	.902	<.001	46.406	<.001	.242	<.001	<.001	-	-	-	-	
		Midair	.815	<.001	.914	.043	.929	.006	14.395	<.001	.004	.401	.004	-	-	-	-	
		Onskin	.901	.023	.924	.073	.905	.028	11.341	.003	.158	.066	.014	-	-	-	-	
Physical Demand	Midair	.801	<.001	.848	<.001	.848	<.001	13.300	.001	.034	.082	.007	-	-	-	-		
	Onskin	.848	<.001	.950	.276	.955	.348	10.173	.006	.043	.200	.013	-	-	-	-		
Temporal Demand	Midair	.921	.060	.899	.020	.958	.408	.955	.343	-	-	-	2.952	.062	.016	-	-	
	Onskin	.899	.020	.958	.240	.964	.515	10.203	.006	.046	1.000	.058	-	-	-	-		
Performance	Midair	.874	.006	.870	.005	.914	.042	1.859	.395	-	-	-	-	-	-	-		
	Onskin	.870	.005	.918	.052	.969	.648	1.126	.569	-	-	-	-	-	-	-		
Effort	Midair	.918	.052	.887	.012	.920	.058	17.816	<.001	.183	.002	.002	-	-	-	-		
	Onskin	.887	.012	.920	.058	.926	.027	20.000	<.001	.009	.091	.002	-	-	-	-		
Frustration	Midair	.799	<.001	.817	<.001	.931	.102	10.265	.006	1.000	.039	.023	-	-	-	-		
	Onskin	.806	<.001	.920	.059	.913	.040	13.618	.001	.113	.043	.018	-	-	-	-		

Table A.2: Statistical comparison of swipe performance metrics across different region segmentation pattern for midair and onskin spaces (RQ2). Figure 6 highlights statistically significant pairwise differences.

B Detailed Results for Section 5.2

This subsection examines how swipe starting and ending regions influence performance in midair and onskin interaction spaces, focusing on the 4- and 6-region segmentations identified as effective in Section 5.1. The 8-region layout is excluded due to limited effectiveness at higher region densities. We analyze both region-level effects of starting and ending region choice and shape-level trends across individual swipe trajectories, highlighting how sensitivity to region selection increases as region density grows and region span narrows.

B.1 4-Region Midair Segmentation

In the 4-region midair condition (Figure B.1), both objective performance and subjective feedback revealed systematic differences among regions. Gesture accuracy (DV1) was significantly influenced by the **starting region** (Friedman $\chi^2(3) = 10.235, p = .017$), but not by the ending region. Swipes initiated from the horizontally aligned regions C (ear/nape, $M = 98.958\%$) and A (nose/eye, $M = 95.833\%$) achieved the highest accuracy overall, with region C significantly outperforming ($p = .027$) region B (temple), which exhibited the lowest accuracy ($M = 88.542\%$). Although accuracy differences across ending regions were not statistically significant, region A exhibited slightly lower ending region accuracy

($M = 89.583\%$), suggesting that swipes terminating near the eye-adjacent region may introduce greater endpoint variability.

Angular displacement (DV4) showed a complementary pattern. Region choice significantly affected angular skew for both swipe initiation (Friedman $\chi^2(3) = 10.012, p = .018$) and termination (Friedman $\chi^2(3) = 19.312, p < .001$). Swipe segments starting from or ending at region A exhibited significantly greater angular skew than those associated with all other regions (A↔B: starting $p = .022$, ending $p < .001$; A↔C: starting $p = .004$, ending $p < .001$; A↔D: starting $p = .013$, ending $p = .002$), with trajectories consistently drifting toward region B. In contrast, swipes performed near the ear and cheek regions (B, C and D) showed substantially smaller angular deviations, indicating more stable midair trajectories.

Subjective **location preference ratings** aligned with these performance trends. Participants reported lower perceived muscle strain and higher comfort for the horizontally aligned regions C and A compared to the vertically oriented regions B (temple) and D (chin), although these differences were not statistically significant. Participants further noted that downward swipes from region B felt more comfortable than upward swipes from region D, reinforcing a general preference for horizontal articulation and downward motion.

Subjective **location preference ratings** did not differ significantly across regions for either initiation ($\chi^2(3) = 4.950, p = .176$)

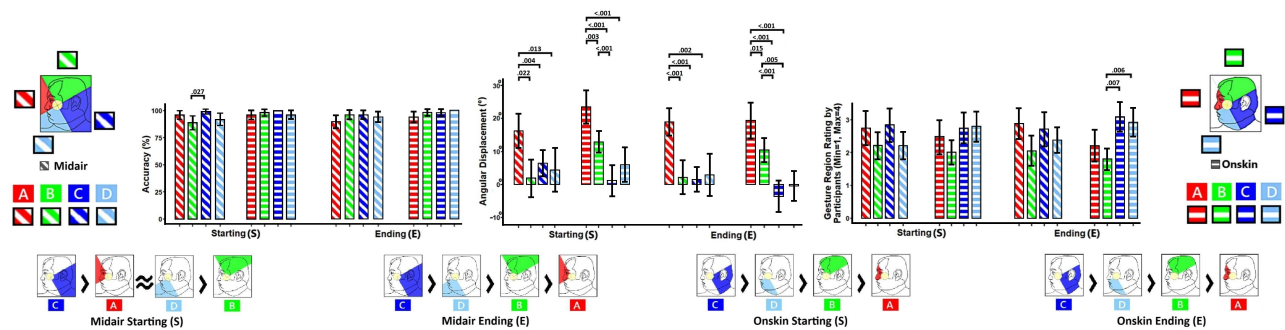


Figure B.1: Accuracy (DV1), average angular displacement (DV4) & location preference ratings for starting and ending regions in 4-region segmentation.

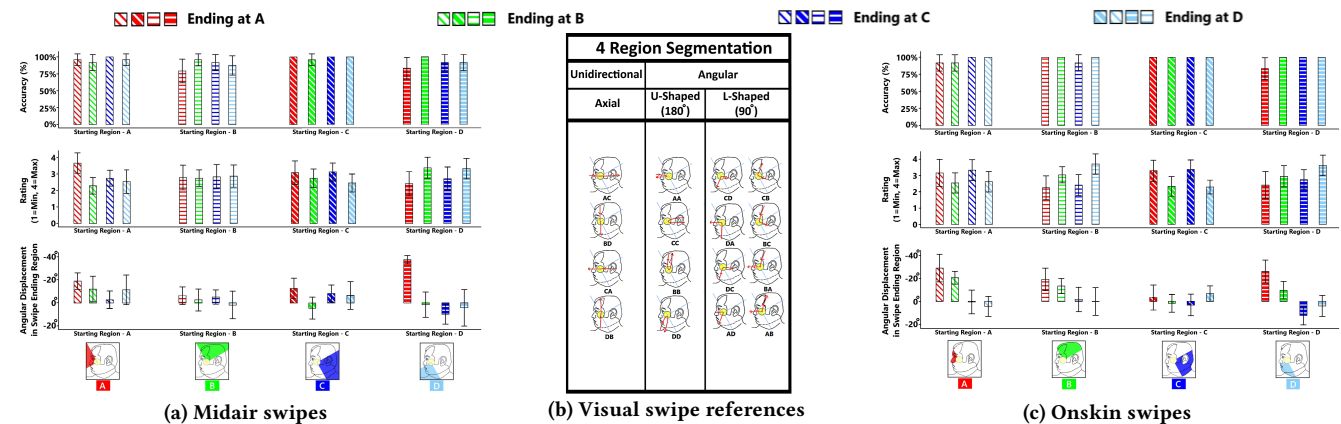


Figure B.2: Accuracy (DV1), average angular displacement (DV4) & location preference ratings for individual swipe shapes in 4-Region Segmentation.

Gesture Metrics	Interaction Space	Position	Shapiro-Wilk Normality Test								Friedman Test for Within Subject Factor		Pairwise Comparison using Wilcoxon Rank-Sum Test			
			Region A		Region B		Region C		Region D		$\chi^2(3)$	p	A	B	C	
			W	p	W	p	W	p	W	p						
Accuracy (DV1)	Midair	S	.20	<.001	.20	<.001	.20	<.001	.20	<.001	10.235	.017	B	.455		
		E	.350	<.001	.200	<.001	.200	<.001	.259	<.001	4.645	.200	B	1.000	.027	
	Onskin	S	.144	<.001	.200	<.001	.127	<.001	.200	<.001	4.400	.221	B	.430		
		E	.259	<.001	.126	<.001	.126	<.001	.144	<.001	8.769	.032	C	1.000	1.000	
		S	.830	<.001	.936	<.001	.973	<.001	.835	<.001	10.012	.018	D	.120	1.000	1.000
		E	.928	<.001	.954	<.001	.969	<.001	.887	<.001	19.312	<.001	B	.022		
Angular Displacement (DV4)	Midair	S	.867	<.001	.939	<.001	.939	<.001	.917	<.001	50.703	<.001	C	<.001	<.001	
		E	.896	<.001	.964	<.001	.923	<.001	.930	<.001	52.088	<.001	D	<.001	.277	1.000
	Onskin	S	.810	<.001	.867	<.001	.807	<.001	.867	<.001	4.950	.176	B	.003		
		E	.831	<.001	.826	<.001	.816	<.001	.882	<.001	5.900	.117	C	<.001	<.001	
		S	.820	<.001	.826	<.001	.836	<.001	.851	<.001	5.750	.124	D	<.001	.007	1.000
		E	.817	<.001	.792	<.001	.793	<.001	.845	<.001	15.850	<.001	B	.190	.006	

Table B.3: Statistical comparison of swipe performance metrics across regions in the 4-region distribution. Figure B.1 highlights statistically significant pairwise differences.

or termination ($\chi^2(3) = 5.900, p = .117$). Nevertheless, descriptive trends suggested slightly higher comfort ratings for regions C and A compared to B and D. Participants further noted that downward swipes from region B felt more comfortable than upward swipes from region D, reinforcing a general preference for horizontal swipes and downward motion.

Taken together, these findings reveal a clear phase-dependent ranking of regions (Figure B.1). For swipe initiation, regions ranked as $C > A \approx D > B$, with region C providing the highest accuracy and stability and region A supporting accurate starts but with greater angular skew. For swipe termination, regions ranked as $C > D > B > A$, with region C again yielding the most stable endpoints and region A performing worst due to increased angular deviation and reduced endpoint control. This asymmetry highlights region C as the most robust choice for both starting and ending midair swipes, and region A as better suited for initiation than termination.

Analysis of individual midair swipe types (Figure B.2a) further revealed shape-dependent performance differences outlined in Table B.4. Overall, unidirectional and U-shaped swipes outperformed most angular L-shaped swipes, particularly in the nose-eye region (AB, BA) and cheek-nose region (AD, DA), which showed the weakest performance among angular variants. Horizontal front-to-back AC swipes along the anterior-posterior axis (from nose

Tier	Gestures	Justification
1 (Best)	DB, CC	Perfect accuracy, strong ratings, and low drift.
2 (Strong)	AC, BB CB, DD	High accuracy with low drift but slightly lower ratings.
3 (Moderate)	BD, BC, DC CD, CA, AA	Good accuracy but moderate drift or ratings.
4 (Weak)	AB, AD, BA	Lower ratings and/or reduced accuracy.
5 (Worst)	DA	High angular drift.

(a) Midair

Tier	Gestures	Justification
1 (Best)	BD, CC, DD	Perfect accuracy, minimal angular deviation, and highest user ratings.
2 (Strong)	AC, CA DB, BB	High accuracy with moderate stability and strong ratings.
3 (Moderate)	AD, CD DC, AA	Reliable but moderate ratings or higher drift.
4 (Weak)	BC, CB	Lower ratings and slightly weaker performance despite stable trajectories.
5 (Worst)	AB, BA DA	Lowest accuracy, highest drift, and lowest rating.

(b) Onskin

Table B.4: Multi-metric swipe ranking in 4-region segmentation based on accuracy, angular displacement, and user ratings.

to ear) and vertical down-to-up DB swipes (from jaw to temple) showed strong overall performance, with DB demonstrating notably low angular displacement and high user ratings. Despite receiving above-average ratings, CA showed significant angular drift at swipe termination, resulting in a lower performance tier. Among U-shaped swipes, the best performance was observed over the ear (CC), followed by the cheek / jawline (DD) and temple (BB) regions. Participants exhibited a preference for U-swipes executed above the nose or along the cheek (CC, DD), despite a slight increase in drift (i.e., angular displacement) near the nose. Participant comments indicated that the spatial extent of regions A and B allowed for minor swipe drift while still remaining perceptually aligned with the intended target area. Visual references for all swipe types are presented in Figure B.2b.

B.2 4-Region Onskin Segmentation

For the 4-region onskin segmentation (Figure B.1), accuracy (DV1) remained consistently high and did not differ significantly for starting regions ($\chi^2(3) = 4.400, p = .221$). For ending regions, however region choice shows a modest but significant effect ($\chi^2(3) = 8.769, p = .032$), with region B exhibiting the highest termination accuracy. Despite this effect, overall accuracy remained above 93% across all regions. Participants reported practical constraints in ocular-adjacent areas - such as surface friction at the temple and hairline (region B) and discomfort or eyewear interference near the eyes (region A) - which were reflected in several low-performing swipe types involving these regions. Consistent with the midair condition, regions distal from the eyes were perceived as more reliable.

In contrast, angular displacement (DV4) varied significantly with facial location and closely mirrored midair patterns (Friedman: starting $\chi^2(3) = 50.703$, ending $\chi^2(3) = 50.703$; both $p < .001$). Gestures involving region A showed the largest angular deviation

(starting $M = 23.448^\circ$, ending $M = 19.375^\circ$), whereas swipe segments performed near the ear and cheek regions (C and D) exhibited the smallest deviations (between -3° to 3° degree from the central plane of the region). Pairwise comparisons confirmed significantly greater deviation for region A relative to other regions (A↔B: starting $p = .003$, ending $p = .015$; A↔C and A↔D: starting/ending $p < .001$). Region B exhibited similar but less severe constraints (B↔C: starting/ending $p < .001$; B↔D: ending $p = .005$). These results indicate that onskin swipes near the ocular area (A and B) introduce greater trajectory variability compared to midair swipes.

Subjective location preference ratings were consistent with onskin accuracy trends, and showed no significant differences for starting regions ($\chi^2(3) = 5.750, p = .124$). However, ending-region ratings differed significantly (Friedman $\chi^2(3) = 15.850, p = 0.001$). Region B (temple) received significantly lower ratings than regions C ($p = 0.007$) and D ($p = 0.006$), indicating reduced perceived comfort and usability near the temple area. Participants frequently reported eye-corner avoidance and hair friction during upward motion as contributing factors to these lower ratings.

Taken together, onskin regions show a clear ranking influenced by anatomical constraints and tactile interaction, with facial contact and ocular proximity imposing stronger limitations than midair swipes. Regions C and D serve as robust anchors, combining lower angular skew with higher comfort and control. In contrast, regions A and B, both near the ocular area, present different constraints. Region B offers a wider interaction area than A, allowing greater steering away from the eye—reflected in reduced skew and more stable accuracy. However, region B is constrained by hair friction and eyewear obstruction, while region A imposes stricter eye-safety constraints, leading to greater angular deviation. Overall, these factors establish our ranking for onskin swipe initiation and termination, as shown in Figure B.1. Regions C emerge as the most reliable anchors for both midair and onskin interactions, with caution advised for precision-critical onskin swipes near the eyes, particularly under eyewear.

Analysis of onskin swipe types reveals shape-dependent performance differences, summarized in Table B.4b and illustrated in Figure B.2b. Unidirectional and most U-shaped swipes exhibit consistently high recognition accuracy, with some achieving perfect accuracy across participants. Among unidirectional swipes, the horizontal temple-to-cheek BD swipe and the swipe between the nose (A) and ear (C) demonstrate reliable performance, with BD showcasing exceptional results, minimal angular displacement, and the highest user ratings.

Gestures near the eye corner display greater directional variability, increasing the risk of accidental ocular interaction. Several L-shaped swipes—specifically AB, BA, and DA—ranked among the lowest in performance due to swipe termination around the ocular area. Among U-shaped swipes, the best performance occurs in the ear (C) and the cheek / chin region (D) (i.e., CC, DD). Temple-centered U-swipes (BB) follow closely, displaying slightly more drift (i.e., angular displacement).

Compared to midair, onskin swipes generally provide better accuracy but exhibit slightly higher drift, affected by tactile contact and the need to avoid ocular interaction. Visual references for all swipe types are available in Figure B.2b.

Gesture Metrics	Interaction Space		Shapiro-Wilk Normality Test												Friedman Test (Within-Subjects)	Post-Hoc Wilcoxon Test with Bonferroni Correction							
	Position	Region	Region A		Region B		Region C		Region D		Region E		Region F			$\chi^2(5)$	p	A	B	C	D	E	
			W	p	W	p	W	p	W	p	W	p	W	p									
Accuracy (DV1)	Midair	S	B	.006																			
			C	1.000	< .001																		
			D	1.000	.572	.778																	
		E	1.000	.023	1.000	1.000																	
		F	1.000	< .001	1.000	.186	1.000																
		G	.329																				
	Onskin	S	B	.043	1.000																		
			C	1.000	.003	< .001																	
			D	1.000	< .001	1.000																	
		E	.007	1.000	1.000	< .001																	
		F	.778	1.000	1.000	.008	.774																
		G	.001																				
Angular Displacement (DV4)	Midair	S	B	< .001																			
			C	< .001	1.000																		
			D	< .001	.570	.910																	
		E	< .001	1.000	1.000	.260																	
		F	< .001	.510	1.000	1.000	1.000																
		G	< .001																				
	Onskin	S	B	< .001																			
			C	< .001	1.000																		
			D	1.000	< .001	< .001																	
		E	.330	.196	.181	.049																	
		F	1.000	< .001	< .001	< .001	1.000																
		G	< .001																				
Location Preference Ratings	Midair	S	B	.860																			
			C	.120	.940																		
			D	1.000	1.000	.200																	
		E	.690	1.000	1.000	1.000																	
		F	1.000	1.000	1.000	1.000	1.000																
		G	1.000																				
	Onskin	S	B	1.000																			
			C	1.000	1.000																		
			D	1.000	.295	.002																	
		E	1.000	1.000	1.000	.016																	
		F	1.000	1.000	1.000	.609	1.000																
		G	1.000																				
Friedman Test (Within-Subjects)	S	B	.223																				
		C	1.000	1.000																			
		D	1.000	.018	.037																		
	E	1.000	.564	.816	1.000																		
	F	1.000	.411	1.000	1.000	1.000																	
	G	1.000																					

Table B.5: Statistical comparison of swipe performance metrics across regions in the 6-region distribution. Figure B.3 highlights statistically significant pairwise differences.

B.3 6-Region Midair Segmentation

Figure B.3 summarizes significant pairwise differences across metrics for both starting and ending regions in the 6-region midair segmentation.

For swipe initiation, starting region significantly affected accuracy (DV1; Friedman $\chi^2(5) = 33.271, p < .001$), angular displacement (DV4; $\chi^2(5) = 96.654, p < .001$), and subjective ratings

($\chi^2(5) = 13.333, p = .020$). Regions **D** (above the ear) and **F** (above the chin) achieved the highest accuracy and lowest angular displacement, indicating stable trajectories during swipe initiation. Regions **A** (above the nose) and **C** (above the temple) followed, exhibiting high accuracy but mild directional bias toward nearby facial landmarks. Region **E** showed reduced performance, likely due to limited visual feedback outside the field of view. The eye-adjacent region **B** exhibited significantly lower accuracy than regions **A** ($p = .006$), **C** ($p < .001$), **E** ($p = .023$), and **F** ($p < .001$). Regions adjacent to eye corner (**B** and **C**) and region around the nape of neck (**E**) showed high angular displacement and ranked among the lowest. Together, accuracy, angular displacement, and subjective ratings revealed a consistent but multidimensional ordering of starting regions, with objective performance favoring $D \approx F > A \approx C > E > B$ (Figure B.3).

For swipe ending regions in the 6-region midair segmentation (Figure B.3), region choice significantly affected swipe accuracy (DV1; Friedman $\chi^2(5) = 34.211, p < .001$), angular displacement (DV4; Friedman $\chi^2(5) = 213.320, p < .001$), and subjective location preference ratings (Friedman $\chi^2(5) = 26.786, p < .001$), indicating that spatial constraints continue to shape swipe trajectories beyond initiation. Region **D** was the most accurate endpoint, outperforming all other regions except **A** ($D \leftrightarrow B p = .003, D \leftrightarrow C p < .001, D \leftrightarrow E p < .001, D \leftrightarrow F p = .008$). **D** also received the highest subjective ratings – significantly higher than **C** ($p = 0.002$)

and **E** ($p = 0.016$)—consistent with its low angular distortion and visually stable termination over the ear. Region **A** followed closely, outperforming regions **C** ($p = .043$) and **E** ($p = .007$). Although endings near the nose (**A**) skewed toward the chin around **F**, angular displacement for **A** remained comparable to regions **D**, **E**, and **F**, supporting both high accuracy and preference. Region **F** showed intermediate performance, with mild skew toward the jawline but sufficient visual feedback for reliable termination. In contrast, regions **C** and **E** produced the lowest accuracy and ratings, as endpoints frequently crossed into adjacent regions ($C \rightarrow D, E \rightarrow F$), reflecting poor spatial separability. Region **B** exhibited highest angular skew—directed toward the temple to avoid ocular interaction. However, due to ocular feedback **B** had slightly better accuracy and ratings than **C** and **E**, although not statistically significant. Together, these measures revealed a clear relative ordering of ending regions ($D > A > F > C \approx E > B$), shown visually in figure B.3.

Individual swipe-shape analysis (Figure B.4a) showed that swipe reliability and preference were strongly shaped by alignment with visually accessible regions exhibiting low angular displacement. Horizontal midair swipes (Figure B.4b) consistently achieved the highest accuracy and preference, followed by U-shaped swipes, indicating that motions aligned with the user’s horizontal field of awareness are more robust than vertical or angular trajectories.

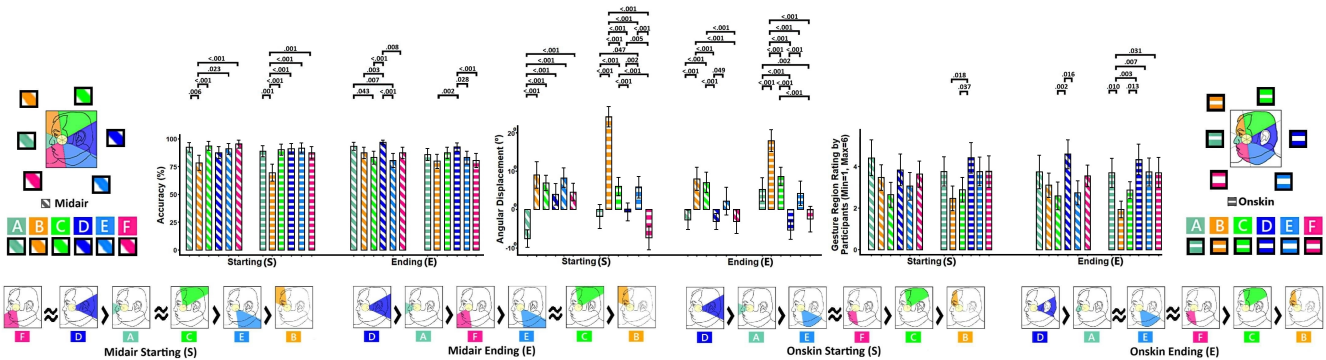


Figure B.3: Accuracy (DV1), average angular displacement (DV4) & location preference ratings for starting and ending regions in 6-region segmentation.

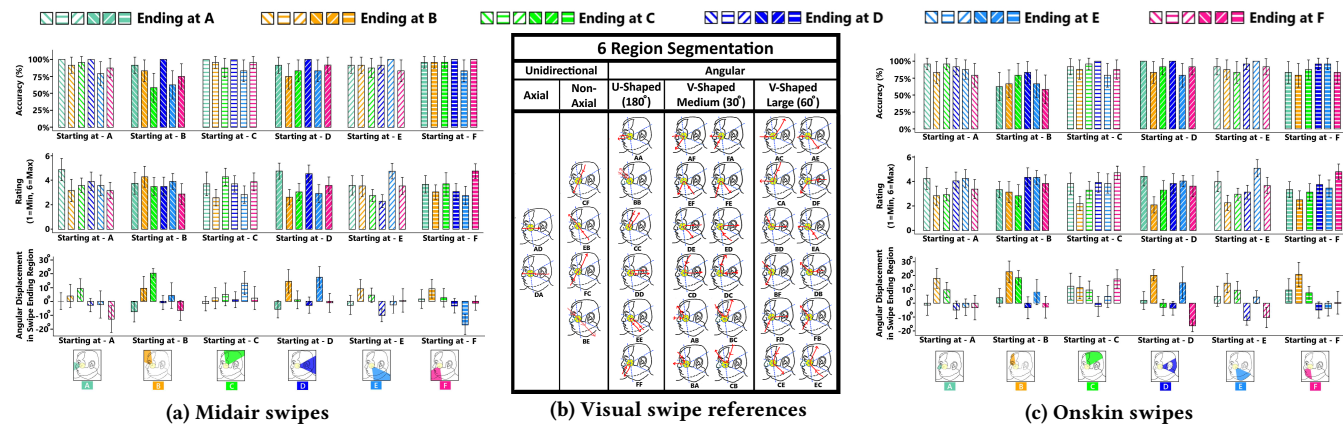


Figure B.4: Accuracy (DV1), average angular displacement (DV4) & location preference ratings for individual swipe shapes in 6-region segmentation.

Non-axial swipes showed mixed performance: swipes directed toward lower-face regions (e.g., **CF**, **FC**) maintained relatively stable trajectories with moderate displacement, whereas swipes involving the eye-adjacent region **B** (e.g., **EB**, **BE**) exhibited greater angular deviation and lower ratings.

For U-shaped swipes, most regions (i.e., **AA**, **DD**, **EE**, and **FF**) remained consistently stable across accuracy and ratings. In contrast, U-shaped swipes in the eye and temple regions (**B** and **C**) showed increased angular deviation and lower ratings, likely due to weaker spatial separation in midair, and reduced visual guidance in the upper-face region - specially for region **C** outside field of view. Among V-shaped swipes, adjacent medium-angle swipes centered near the ear-adjacent region **D** (e.g., **CD**, **DC**, **FA**) maintained strong accuracy with moderate displacement, whereas swipes intersecting upper-face regions (e.g., **AB**, **BA**, **BC**, **CB**) showed greater angular deviation and reduced stability. Similarly, large V-shaped swipes spanning non-adjacent regions remained relatively robust when aligned along the posterior cheek axis (e.g., **CA**, **DF**, **EA**), but degraded when trajectories crossed upper-face regions (e.g., **BF**, **DB**, **CE**). Midair swipes **CE** and **EC**, which span visually constrained regions, frequently drifted toward region **D**, resulting in lower accuracy and preference; these swipes therefore align more closely with **DD** than with distinct region-to-region inputs.

Overall, midair swipe patterns show that executing swipes in upper-face regions **B** and **C** is challenging due to visual avoidance and spatial drift. Drift pattern observations suggest merging **BB**, **BC**, and **CC** into a U-shaped midair swipe (temple–cheekbone–temple), which lowers cognitive load while maintaining expressive interaction. Additionally, the drift from **F** to **E** for **FE** swipe indicates that **FE** and **EF** could also combine into a single midair swipe, aligning with the U-shaped **EE** swipe. Furthermore, downward swipes from **B** / **C** to **E** / **F** (i.e., **CE**, **BF**, **BE**, **CF**) can be unified as a swipe-down action, while corresponding upward swipes (**EC**, **FB**, **EB**, **FC**) can be unified to create a complementary swipe-up action.

B.4 6-Region Onskin Segmentation

Figure B.3 summarizes all significant pairwise differences across metrics for both starting and ending regions in the 6-region onskin segmentation, with post-hoc p -values annotated. For swipe initiation, region choice significantly influenced angular displacement (**DV4**; Friedman $\chi^2(5) = 212.310$, $p < .001$), which in turn explained differences in accuracy (**DV1**; Friedman $\chi^2(5) = 52.981$, $p < .001$) and starting-region self-ratings (Friedman $\chi^2(5) = 17.167$, $p = .020$). The eye-adjacent region **B** (around the eyebrow and outer eye corner) consistently performed worst, exhibiting significantly lower accuracy than all other regions ($p < .001$) and substantially higher angular displacement ($p < .001$). Because region **B** occupies a small area tightly surrounding the eye in the 6-region layout, swipe segments frequently deflected into the temporal region **C**, increasing skew and retries. Region **C** (temple) showed improved accuracy relative to **B**, but remained less preferred, as reflected in its lower ratings.

In contrast, region **D** (cheekbone–tragus junction) emerged as the most effective and preferred starting location, receiving significantly higher ratings than regions **B** ($p = .018$) and **C** ($p = 0.037$)

and exhibiting among the lowest angular displacement values. Region **A** (around the nose) showed comparable accuracy and angular stability to **D**, with no statistically significant differences between them. Regions **A** and **F** (mouth / chin) benefited from combined haptic and visual feedback, supporting more controlled initiation and reduced overshoot. Regions **E** (nape) and **F** showed similar objective performance, although **E** was rated slightly lower due to increased wrist bending. Together, these results support the ordering $D > A > E \approx F > C > B$ for onskin swipe initiation (Figure B.3).

For swipe termination, ending-region choice similarly affected accuracy (**DV1**; Friedman $\chi^2(5) = 23.073$, $p < .001$), angular displacement (**DV4**; Friedman $\chi^2(5) = 147.550$, $p < .001$), and ending-region self-ratings (Friedman $\chi^2(5) = 26.786$, $p < .001$), indicating persistent region-dependent effects, although termination was more tolerant of brief traversal through constrained areas. Region **D** again yielded the most reliable endpoints, achieving higher accuracy than regions **B** ($p = .002$), **E** ($p = .028$), and **F** ($p < .001$), and receiving the highest ratings, significantly exceeding those of regions **B** ($p = .003$) and **C** ($p = 0.013$). Region **B** ranked lowest overall, exhibiting significantly greater angular displacement than all other regions ($p < .001$) as endpoints deflected toward the temple to avoid ocular contact, resulting in the lowest accuracy and ratings. Region **C** showed reduced skew and improved accuracy relative to **B** ($p < .001$), but remained suboptimal due to limited endpoint separability. Regions **A** and **D** exhibited comparably low angular displacement and high accuracy, while region **F** showed intermediate performance; region **E** was consistently less preferred. Across phases, the ordering $D > A \approx E \approx F > C > B$ was preserved for onskin swipe termination (Figure B.3).

Swipe-shape analysis (Figure B.4c) revealed systematic swipe stability patterns across the 5 swipe groups summarized in Table 4b. **Axial swipes** (i.e., **AD**, **DA**) were highly stable, achieving high accuracy, low angular displacement, and favorable ratings due to continuous tactile guidance along the cheek. In contrast, **non-axial swipes** showed mixed performance: swipes confined to cheek regions (e.g., **CF**, **FC**) maintained relatively high accuracy with moderate displacement, whereas swipes involving the eye-adjacent region **B** (e.g., **EB**, **BE**) exhibited greater angular deviation and lower ratings. For **U-shaped swipes**, regions **A**, **D**, **E**, and **F** remained consistently stable across DV1 and ratings and outperformed vertical non-axial swipes. In contrast, **BB** and **CC** showed increased angular deviation and lower ratings, reflecting upper-face interaction constraints: region **B** lies adjacent to the eye, where users tend to avoid contact with the eye corner and drift towards **C**, while region **C** may introduce additional variability due to hair-related friction.

Among **V-shaped swipes**, adjacent medium-angle swipes avoiding region **B** (e.g., **CD**, **DC**, **EF**, **FE**, **FA**) maintained high accuracy with moderate displacement, whereas swipes intersecting the eye region (e.g., **AB**, **BA**, **BC**) showed greater displacement and lower ratings. Similarly, **V-shaped large swipes** spanning non-adjacent regions remained robust when anchored on the cheek (e.g., **DF**, **FD**), but degraded when involving region **B**. Regions **E** and **F** also exhibited distinct swipe behaviors, likely reflecting different tactile anchors along the mandible—region **E** near the mandibular angle and region **F** near the anterior jaw and chin—suggesting that these lower-face regions should remain separate.

Overall, onskin swipe drift patterns indicate that swipes originating or terminating in regions **B** and **C** remain challenging for users, particularly near the eye and temple areas. Merging **EC** with **FC** and **CE** with **CF** may therefore improve recognition robustness for vertical swipes while reducing reliance on the eye-adjacent region.

More broadly, these patterns suggest that robust onskin swipe sets in the 6-region layout should prioritize horizontal, cheek-anchored trajectories while minimizing interaction near the eye and temple regions.