# Exploring the use of hand-to-face input for interacting with head-worn displays

Marcos Serrano, Barrett Ens, Pourang Irani

# Exploring the Use of Hand-To-Face Input for Interacting with Head-Worn Displays

**Marcos Serrano[1,2], Barrett Ens[1], Pourang Irani[1]**
[1]University of Manitoba
Winnipeg, MB, Canada, R3T 2N2
{marcos, bens, irani}@cs.umanitoba.ca
[2] IRIT - University of Toulouse
Toulouse, France, 31062
marcos.serrano@irit.fr

## ABSTRACT
We propose the use of *Hand-to-Face* input, a method to interact with head-worn displays (HWDs) that involves contact with the face. We explore Hand-to-Face interaction to find suitable techniques for common mobile tasks. We evaluate this form of interaction with document navigation tasks and examine its social acceptability. In a first study, users identify the cheek and forehead as predominant areas for interaction and agree on gestures for tasks involving continuous input, such as document navigation. These results guide the design of several Hand-to-Face navigation techniques and reveal that gestures performed on the cheek are more efficient and less tiring than interactions directly on the HWD. Initial results on the social acceptability of Hand-to-Face input allow us to further refine our design choices, and reveal unforeseen results: some gestures are considered culturally inappropriate and gender plays a role in selection of specific Hand-to-Face interactions. From our overall results, we provide a set of guidelines for developing effective Hand-to-Face interaction techniques.

### Author Keywords
HMD; HWD; Head-worn display; mobile interfaces; body interaction; input techniques

### ACM Classification Keywords
H.5.2. Information interfaces and presentation: Interaction.

## INTRODUCTION
Head-mounted devices are becoming available for widespread, daily use through lighter form factors and with transparent displays. We refer to these modern accessories as head-worn displays (HWDs). As consumers may soon get affordable access to HWDs [6, 30, 31], ways in which they interact with content on such devices is being actively investigated [17].

Currently, HWDs provide onboard microphones and small capacitive sensors for user input. Voice recognition is useful for command-based tasks such as for search queries but has limited use in certain settings (i.e. noisy environments). The capacitive surface on the temple of HWDs presents a viable on-device method for input, but it has limited input space. Other less self-contained options, such as a wearable device or an auxiliary smartphone, can also allow for input [21,26]. However, these require carrying and may be occluded by the HWD content. Natural user interfaces [29] can overcome the above limitations. However, mid-air input [2,14] suffers from the lack of tactile feedback and on-body gestures [8, 9] such as making contact with the arm skin [8], are often coupled with on-body projection for output.

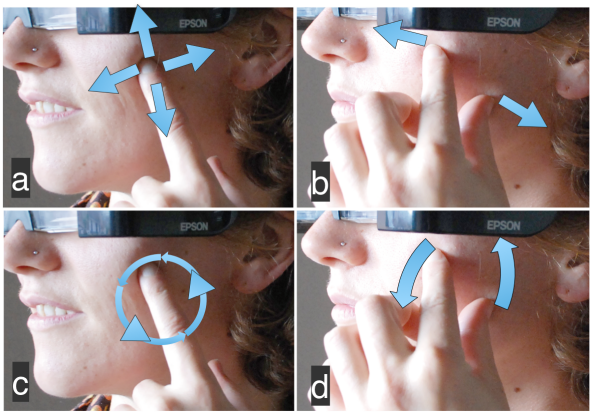


**Figure 1. Hand-To-Face input for navigation. a) Panning, b) Pinch zooming, c) Cyclo zooming, d) Rotation zooming. Our studies show that Cyclo was not socially acceptable while Rotation was not efficient.**

We propose *hand-to-face input* as a novel, alternative method for interacting with HWDs. We define hand-to-face input as any gesture that involves touching, rubbing, scratching or caressing the face. This approach is especially well-suited for interaction with HWDs for many compelling reasons: (i) the face is often touched [18, 20] making it a promising area for subtle interactions; (ii) it offers a relatively large surface area for interaction, but not normally clothed as are other areas; (iii) it facilitates eyes-free, single-handed input, which can be invaluable in mobile settings (e.g. riding a bike, holding on in a bus); and (iv) is in close proximity to the HWD, making it likely to accommodate device-borne sensors and creating a natural extension of the device temple.

We first explore the design space of hand-to-face input by eliciting from users the range of gestures for various mobile tasks, such as navigation and action selection. Our study

M. Serrano, B. Ens and P. Irani. Exploring the use of Hand-To-Face Input for Interacting with Head-Worn Displays. In CHI'14: *Proceedings of the SIGCHI Conference on Human Factors and Computing Systems*, 10 pages, to appear, ACM, 2014.

participants generally identified the cheeks and forehead as good surfaces for gestures. Based on these results, we designed hand-to-face navigation techniques (Figure 1) and found these to be most effective and least tiresome when carried out on the cheek. Given this non-ordinary form of interaction, we also examined whether hand-to-face interaction was perceived to be socially suitable. In most cases, participants found this form of input acceptable.

Our contributions include 1) an elicitation of potential hand-to-face gestures for mobile tasks with HWDs; 2) a design of hand-to-face input techniques for document navigation; and, 3) a validation of the suitability of such interactions for use in public settings.

## STATE OF THE ART

Our research is inspired by recent results on interaction methods for wearable displays, in particular work on mid-air gestures and on-body input. We also discuss the implications of casually touching the face in daily settings.

### Mid-air gestures

Mid-air gestures have gained significant interest recently due to the development of readily available gesture tracking systems. Mid-air gestures, using one or two hands, often require additional equipment, for example on the shoes [2], or on the wrist [14] to capture user input. Other forms of gestures such as foot input [1] or head movement [28], are possible but would affect viewing stability on a HWD. For the most part, mid-air gestural input suffers from a lack of haptic feedback.

### On-body interaction

A large body of work has considered appropriating the body as an interaction surface. On-body interaction offers an input surface, the human skin, with the advantage of leveraging human proprioception as an extra feedback mechanism. This can overcome some of the limitations with mid-air interactions. Much of this prior work has considered coupling on-body input with on-body projection using wearable pico-projectors, such as with Skinput [8], Brainy Hand [29] and OmniTouch [9]. These projection-based techniques are also adaptable to HWDs.

Gustafson et al. [7] investigated using palm-based imaginary interfaces, i.e. interfaces without any visual feedback. Their results show that tactile sensing on the palm allowed users to orient themselves more effectively than cues sensed by the pointing finger. Likewise for hand-to-face interaction, cues sensed by the facial skin could actually help orient the user. Recent work by Wagner et al. [32] proposed a body-centric design space to describe multi-surface and on-body interaction. This study however omitted exploring interactions on the face.

Only a few studies have explored the use of specific face areas, such as the ear [16] or the tongue [25] for input. We explore the range of different face areas, focusing on the design and social suitability of hand-to-face input, as well as its impact on effective document navigation.

### Hand-to-face casual gestures

Hand to face contact is frequent and can lead to an average of 15.7 contacts/hour in work settings [20]. As contact with the mouth, eyes and/or nostrils can lead to the transmission of diseases, hand-to-face input should not include such health-sensitive areas. More recently, researchers studied hand-over-face postures in communication involving different emotions (frustration, surprise…): they found that 13% of the gestures were on the upper vs. 89% on the lower regions of the face (some covered both regions) [18]. These results further suggest that implicit face contact can be subtle, frequent and natural. We seek to make face contact explicit for common mobile tasks.

## DESIGNING HAND-TO-FACE INTERACTIONS

Factors influencing the design of hand-to-face interaction include: facial area of interaction, number of fingers used, gesture style, physical demand, social acceptability and distinguishing explicit from implicit input

### Area of interaction

Prior research suggests which facial areas are commonly used (lower regions [18]), as well as those to avoid (mouth or eyes [20]). In this first exploration, we investigate the range of gestures possible on defined areas of the frontal part of the face, such as the chin, cheek or forehead. We explore the feasibility of input on areas of the HWD such as the temple, hinge, bridge, frame and glass. We identify which regions are more effective and less prone to fatigue.

### Number of fingers

The use of multiple fingers can affect hand-to-face input as certain regions on the face or the wearable display have limited surface area. However, using multiple fingers can be natural for mobile tasks such as zooming in/out of a map. HWDs offer a rather limited surface (Figure 2), which can make multi-finger interaction difficult. The face seems more appropriate for this type of interaction due to its large dimensions. We study this aspect in our second experiment.

### 1D vs. 2D Gestures

The temple of several contending HWDs consist of a long, narrow surface (Figure 2). As a result, 1D gestures, such as flicking are better suited than 2D ones, such as pinching. Moreover, we expect large areas of the face such as the cheek to be better at supporting 2D gestures, such as panning on a map. We elicit from study participants potential gestures for 1D and 2D tasks.

### Physical demand

Lifting the arm to touch parts of the face, such as the neck or the chin, can be physically demanding. However, touching the upper areas of the face, such as the hair or the forehead, requires more effort as the user has to perform larger arm movements. We investigate the physical demand of interacting with different areas of the face in our studies.

### Social acceptability

Since casual hand-to-face gestures are common, they can be perceived as a non-intrusive and subtle input method.

Conversely, users may be reluctant to perform on-body interaction, particularly on areas that garner social attention, such as the face. It could also interfere with users' facial cosmetic products. To address these issues, we examine the social acceptability of such input after identifying the most effective hand-to-face techniques.

**Figure 2. HWDs ordered by increasing temple size: Telepathy [30], Google Glass [6], Vuzix M100 [31] and Moverio [4].**

### Implicit and Explicit Gestures

Given the frequency at which face contact occurs, hand-to-face gestures need to be invoked explicitly to avoid the Midas touch problem [12], i.e. inadvertently issuing a command to the device. This can be avoided using a delimiter at the start of the input, such as through voice (e.g. "HWD-navigate"), by pressing a button on the display or by pressing harder for explicit input. This concern is not unique to face input, as interacting with HWDs also requires that the user enter a specific mode. We leave out the examination of this factor in this first exploration.

## GUESSABILITY STUDY

To explore the breadth of potential hand-to-face gestures and their mapping to interactive tasks, we elicited user input through a guessability study [1,24,27,33]. For a set of common mobile tasks, as in [24], we asked participants to suggest suitable gestures on the face (above the neck) and on the HWD.

### Overview and rationale

For exploring potentially rich and vast gesture sets, user-elicitation or guessability studies have shown favourable results [1, 24, 27, 33]. Wobbrock et al. [33] found that eliciting gestures from users resulted in over 40% more gestures than if asked by expert designers. This motivated the use of such an approach to identify gestures for a multi-touch tabletop [33], for mobile motion gestures [24], and for foot interaction [1]. As in these previous works, we focus primarily on human capabilities. Consequently, we put aside any recognizer issues and asked users to perform gestures at their will without worrying about the underlying sensing technology. We asked participants to include gestures on the entire face, i.e. any region on or above the neck. This allows for a larger set of potential gestures. We also assessed users' preference for interacting with either the face or areas of the HWD, for each of the tasks.

### Mobile tasks considered

Most HWDs possess similar applications to those currently available on smartphones. For example, demonstrations of Google Glass [6] show users taking photos and videos with a camera, navigating a map, texting, selecting phone calls and even browsing the internet [6]. In some cases, HWDs are designed as an auxiliary companion to mobile devices, such as the Vuzix M100, a wearable wireless display for Android smartphones [31]. For these reasons in this first exploration, we elicit gestures for common mobile tasks. Inspired by Ruiz et al. [24] we divide tasks into two categories: action (tasks 1 to 9) and navigation (tasks 10 to 22). Action includes tasks such as answering a call, taking a picture or selection. Navigation includes tasks such as opening the home screen, discrete 2D navigation or continuous 2D panning and zooming.

### Participants

Fourteen (6 female) volunteers participated, of 27.8 years on average. We did not filter out participants who had experience with natural user interfaces. As a result, of our participants 7 were familiar with mid-air gestures, 1 had seen on-body interaction (in a video) and one had used a HWD. Users were all unfamiliar with our proposed hand-to-face input. We rewarded them with a gift card.

### Apparatus

We mocked up a HWD using a plastic set of sunglasses (Fig. 3) with a larger craft board temple as the interactive surface. To define the size of the interactive temple we calculated the average width of that used on five commercial HWDs. The size of this interactive temple was 11×2 cms. The weight of the added piece was negligible.

**Figure 3. Mockup of the HWD used in the study.**

### Procedure

We familiarized our participants with HWD capabilities by showing them a demonstration video of Google Glass [6]. The video shows a first-person view of the display features of the device (such as image or map browsing) without showing the interaction modalities. We showed this particular video to not bias participants to current HWD interactions. We asked participants to design gestures for a HWD for each of the given tasks. Participants had to generate one gesture for hand-to-face input and another for use on any part of the HWD surface. Unconstrained by any time limit, users wrote down the most suitable gesture for each task on a schematic front and profile face image (similar to Figure 4). After sketching the gestures for a set of tasks, we asked them to perform their gestures, which we video recorded. Finally, we asked users to select and justify, for each task, which gesture they preferred, i.e. the one on the HWD or on the face. This took about 60 minutes in total. Participants wore the mock HWD for both conditions.

### Collected data

Every user generated one sheet with a sketch of the gesture on the face and on the display for every task. We video recorded the entire experiment and users' verbal comments.

## RESULTS

We collected gestures for 22 tasks × 14 users × 2 conditions (face and HWD) for a total for 616 gestures. The complete

gesture set is available online (http://hci.cs.umanitoba.ca/face-input).

### Comparing interaction for the face and the HWD

*Areas of interaction*

On the face, participants produced gestures for a total of 11 different areas, such as the cheek, forehead, ear, chin or jaw. The results reveal a distribution (Figure 4) with gestures concentrated on the cheek (34%) and then on the forehead (16%). Other areas saw an equal distribution of gestures: jaw (8.7%), ear (7.7%), temple (7.4%), and chin (7.4%). Areas such as the eyes, nose, lips, neck and hair were least used. On the HWD, participants used 5 different interaction areas: temple, hinge, frame, bridge and glass. Most of the gestures (60%) were situated on the temple.

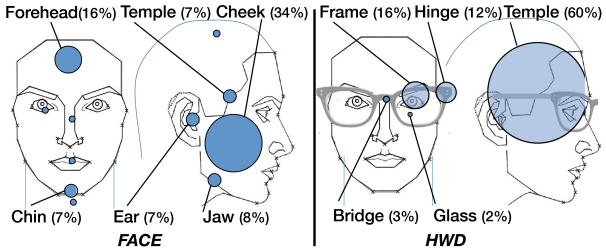


**Figure 4. Main areas identified by participants as suitable for input on the face (left) and on the HWD (right). The circles' size is proportional to the percentage of gestures (in brackets).**

*User preference*

Overall users preferred interaction on the face for the *navigation* tasks (T10-T22), while opinions were mixed for the *actions* tasks (T1-T9) (Figure 5). Users particularly preferred using the face for panning (10/14 preferred the face) and zooming (9/14). Users indicated that *"the face provides a larger area"* [P2], which is perceived as a benefit for panning and zooming. This is particularly true when using the cheek, since it is "*the best part to interact with the face*" [P9] and it is "*like a touchpad*" [P1].

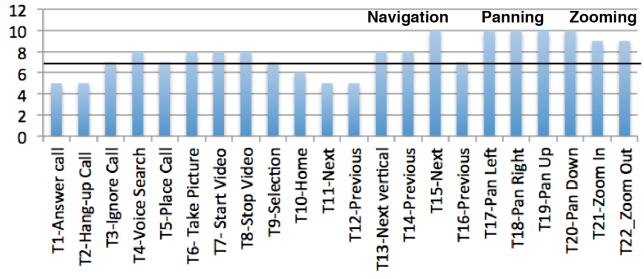


**Figure 5. Number of users preferring face to HWD for each task. We emphasize the number 7 (50% of the 14 users).**

*Benefits and limitations of face and HWD input*

We summarize users' perceived benefits and limitations for interacting on the face and the HWD in Table 1. These results are influenced by the form factor of the mockup used in the study, whose general characteristics (weight, inertia or grip on head) differ from popular HWDs. For instance, physical discomfort could be improved in a commercial HWD, although continuous input on the temple may still be irritating. Interestingly, two female users preferred the HWD for all tasks; one suggested that face input would interfere with facial cosmetics and the other felt that it would be socially awkward. We examine these issues in our final study.

| HWD | Face |
|---|---|
| + ***Metaphor:*** "for some tasks, the HWD make sense, such as taking a picture, it's like a camera" [P2] | + ***Large surface:*** "the face has a larger surface" [P14] |
| + ***Rapid activation:*** "HWD is useful for single taps and for smaller gestures" [P5] | + ***Natural interaction:*** "the face makes you feel more natural" [P2] |
| + ***Physical gestures:*** "holding and moving the HWD is nice" [P9] | + ***Symbolic gestures:*** "mapping gestures is easy, for instance tapping near the ear" [P3], |
| – ***Limited space:*** "HWDs have limited space, specially for panning and zooming" [P9] | + ***Memorization:*** "Face is more meaningful, easier to remember" [P8] |
| – ***Physical demand:*** "HWDs are higher up on the face" [P8] | + ***Flat surfaces:*** "cheek is like a touchpad" [P1], "forehead is flat and colinear to the field of view of the HWDs" [P4] |
| – ***Instability:*** "HWDs move when touched" [P7], "HWDs shake" [P6], "touching the HWD may move the camera" [P10] | – ***Makeup:*** "did not want to touch the face due to makeup" [P11] |
| – ***Occluded vision:*** "occlusion of field of view when using the HWDs" [P4] | – ***Dirt:*** "finger skin oil will remain on forehead" [P12] |
| – ***Physical discomfort:*** "tapping the HWD hurts my nose" [P9]. | – ***Sensitive surface:*** "the face is too sensitive, touching often may hurt" [P3] |

**Table 1. User comments on benefits (+) and limitations (-) of HWD and Face interaction.**

### Participant agreement on hand-to-face gestures

Participants produced a large variety of hand-to-face gestures (Figure 6) which we analyze in two steps. We use prior methods [33] for the first analysis while the second one is based on a taxonomy we infer from users' responses.


**Figure 6. Examples of gestures identified during the study by participants. a) Touching the mouth to place a call; b) touching the hair to ignore a call; c) rotating two fingers, or d) touching the chin to do a voice search; e) swiping the forehead to start video recording; f) pinching the frame to take a picture (on the HWD); g) tapping on the HWD bridge to go to the home screen; and h) grabbing the ear to ignore a call.**

*First analysis*

We analysed the agreement between participants for the set of gestures produced for each task using Wobbrock's approach [33]. The agreement value ranges between 0 (no

agreement) and 1 (total agreement) and indicates whether users agreed on using a specific gesture for a given task. We group gestures which are of the same type (swipe, tap, etc.) and occur in the same area (cheek, chin, etc.).

The mean value for the agreement score is 0.14 (SD 0.06), with 36% of the tasks having an agreement value higher than 0.2% (Figure 7). While this score seems low, it is on par with that from other previous guessability studies [1]. The gestures for panning (0.28) showed the higher agreement score. The two main panning gestures were "cheek swipe" (Fig. 1-a) and "forehead swipe" (same but on forehead). The gestures for zooming (the others preferred by users) have an agreement score of 0.13.

This low agreement score is mainly due to the variety of areas used for zooming (ear, hair, nose, neck, mouth, jaw, forehead, chin and cheek). Swiping the cheek is considered different than swiping the forehead. We propose a taxonomy to describe hand-to-face gesture properties to evaluate gesture similarity in our second analysis

*Hand-to-face gestures taxonomy*

Through an open coding analysis, we infer a hand-to-face gesture taxonomy from the results of the study above, from the state-of-the-art on implicit hand-to-face gestures [18] and from previous guessability studies on motion gestures [24]. While this taxonomy could include many properties, such as the number of hands, we describe only the properties sufficient to characterize our results. Our taxonomy includes five properties to describe the gesture mapping and its physical characteristics: **1) Nature:** the nature of the gesture can be a *metaphor* (scratching the face or covering the mouth, 56% of total), *physical* (grabbing the ear, 4%), *symbolic* (drawing a letter or a symbol such as a circle, 2%) or *abstract (38%)*. **2) Temporal:** the gesture can be *discrete* (48%) or *continuous (52%)*. **3) Pose:** the gesture can be a static (36%) or a dynamic pose (64%). **4) Number of fingers:** gestures can use one (79%), two (17%) or several fingers (4%). **5) Area:** we divide the face into the following areas: cheek, forehead, temple, ear, eye, chin, mouth, neck, jaw, hair and nose (see percentages on Fig. 4).

For instance, the gesture *i*n Figure 6-a, consisting of touching the mouth to place a call, can be described as *<metaphor, discrete, static, two-fingers, mouth>*.

*Second analysis*

Based on the previous taxonomy, we define a formula to calculate the similarity score, which indicates whether different gestures share common properties. The similarity score $S_t$ of a task *t* is the average of the agreement for every property $P_i$ of our taxonomy, from the set of properties Pt. To calculate this value we use the formula of the agreement score [33]: $G_i$ is the subset of gestures with identical value for the property $P_i$ from the set of gestures $G_t$.

$$St = \frac{\sum_{Pi}\left(\sum_{Gi}\left(\frac{Gi}{Gt}\right)2\right)}{Pt}$$

The mean value for the overall similarity score (Figure 7) is 0.61 (SD 0.1) with panning (0.73) and zooming (0.72) having the highest agreement scores. The *Nature* of zooming gestures (0.86) is mainly based on the metaphor of pinching with two fingers. The *Pose* for zooming is always dynamic (1.0) and most subjects used two fingers for zooming (0.80).

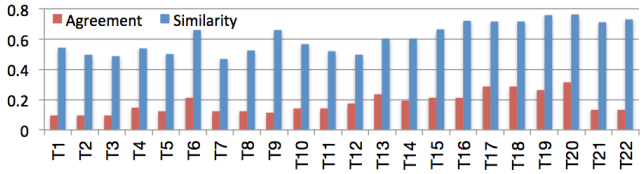


**Figure 7. Agreement and similarity scores of hand-to-face gestures for each task. Panning (T17-T20) shows the highest overall scores while zooming (T21-T22) has a high similarity.**

### Summary

The guessability study reveals that users find hand-to-face gestures appealing for navigation tasks (panning and zooming). We can describe panning gestures using our taxonomy as *<abstract, continuous, dynamic, one finger, cheek/forehead>* and zooming gestures as *<abstract, discrete/continuous, dynamic, two-fingers, cheek/chin/forehead>*. Given the high prevalence for document navigation in mobile contexts (a recent survey reveals Google Maps is the most used mobile application worldwide [5]), we further investigate the performance, fatigue and user preference of hand-to-face input for panning and zooming.

## HAND-TO-FACE NAVIGATION STUDY 1: PANNING

In this study we evaluate the performance, physical effort and user preference of hand-to-face gestures for panning.

### Participants

Twelve students (2 female) from a local university volunteered with an average of 25.9 years and were rewarded with a gift card.

### Apparatus

We used the Epson Moverio BT-100 a commercially available HWD. The device weighs 220g, has a resolution of 960×540 pixels and a refresh rate of 60Hz. The device measures 17.8×20.5×4.7cm and runs Android 2.2. We included two different temple sizes (Size factor). The Oversized temple corresponds to the actual device's temple, which measures 7×5cm (larger than most HWDs to be released). The Regular temple is based on the average temple size used in our guessability study, measuring 7×2cm. We used craft board to mock up the temples.

We implemented face gestures using a Vicon T20 infrared optical tracking system with six cameras positioned around the user (front, front-right and right side at different heights). We placed infrared markers on the participant's index finger (Figure 8). To detect skin contact, we used a proximity sensor connected to a micro-controller. The sensor set was connected through a USB cable to a desktop computer. To make the sensor set unobtrusive, we

integrated it into a glove worn by the user. During our studies we had a negligible number of tracking errors (0.15% of all trials). The system had no perceivable latency: all input was merged to the same program and sent to the HWD through USB. Optical tracking ran at 690 Hz; contact sensor at 600 Hz; merged data were sent to the HWD at a measured rate of 142 Hz.

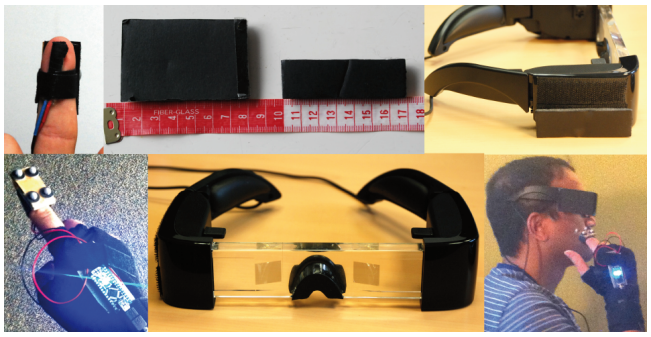

**Figure 8. Top: Sensor used to detect skin contact; Oversized and Regular temples measures (cms); Regular temple attached to HWD. Bottom: IR markers and microcontroller on hand; Epson Moverio HWD and subject wearing HWD.**

### Task and mappings

Users were asked to pan until the target, represented by the smallest central circle in a set of concentric rings, is in the screen center (Figure 9). The target is placed in one of 6 different directions (2 vertical, 2 horizontal and 2 diagonals), at 3 screen widths distance. Participants used one of four different interactive surfaces, two on the face (cheek and forehead) and two on the display (Oversized and Regular temple). We include two different temple sizes in order to study the impact of its size on navigation.

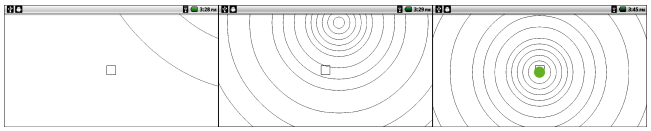

**Figure 9. Visual feedback of the panning task displayed on the HWD from initial state (left) to target reached (right).**

We use three panning interactions: displacement-based (D), flick (F) and rate-based (R). With Pan-D, finger movement is directly mapped to the movement of the map. In Pan-F, the user flicks to pan, mimicking the iOS flick behaviour [10]. In Pan-R, the distance panned from the initial touch position is mapped to the finger velocity movement.

### Design

The experiment used a 4×3×3 design with Area (Cheek, Forehead, Oversized temple, Regular temple), Gesture Mapping (Pan-D, Pan-F, Pan-R), and Direction (Horizontal, Vertical and Diagonal) as factors. The experiment is divided into 12 blocks, each corresponding to a specific area. Each block was repeated three times. Order of blocks is counterbalanced across participants. Trials in the block were grouped by mapping (i.e. all trials of flicking were performed together) always in the same order (Pan-D, then Pan-F, then Pan-R). Every direction is performed two times per block. Every condition is repeated 6 times in total.

### Procedure

Participants were allowed to get used to the apparatus and were given sufficient training, first on the desktop display, then on the HWD. The experiment was divided into blocks. The user starts a block by pressing a key. In each block, there is a 2 sec interval between trials. We informed users that they can take a break between blocks and ask them to be as fast as possible during the trials. Participants wore the HWD during the entire experiment.

### Collected data

We logged all tracking data and measured time to complete the task from stimulus onset and first touch. We measured physical demand by using the Borg CR10 scale of perceived exertion, specifically adapted to physical demand [3,11]. We also measured user preference using a 5-point Likert scale to rate the four areas and the three techniques. We collected 216 trials per user (4 areas × 3 techniques × 3 directions × 6 repetitions) × 12 users = 2592 trials in total.

### Results

*Time Performance*

Trial completion time (Figure 10-left) which was measured from first touch until target reached, we found a significant effect of Area ($F_{3,33}$ = 12.7, **p**<.0001), Technique ($F_{2,22}$ = 52.7, **p**<.0001) and Orientation ($F_{2,22}$ = 13.9, **p**<.0001). There is an interaction effect between Area and Technique ($F_{6,66}$ = 8.2, **p**<.0001) and between Technique and Orientation ($F_{4,44}$ = 7.1, **p**<.0001). The interaction effect between Area and Technique is mostly due to the significant time taken to complete the task with the Pan-R (Rate) technique on the regular temple. In this particular case users' finger often accidentally slid out of the temple due to its small size, forcing users to clutch often.

For Area, a post-hoc test reveals that Cheek (mean 6.5s) is significantly faster than Forehead (mean 8.2s) and Regular Temple (mean 8.9s). We found no significant difference between Cheek and Oversized Temple (mean 7s). This can be expected as they have a similar effective surface: the cheek is generally larger but participants tended to use its central area. With Technique, a post-hoc test reveals that Rate (mean 9.8s) is significantly slower than Displacement (mean 6.6s) and Flick (mean 6.4s). While Rate was particularly slow for the Regular temple, it also showed worst performance on Cheek and Oversized temple. As expected, Horizontal Orientation (mean 6.5s) is significantly faster than Vertical (8.1s) and Diagonal (8.3s).

*Fatigue and User preference*

To analyze the results from the Borg questionnaire (Figure 10-right) we perform an Anova test (Borg being a scalar value). There is a significant effect of both Area ($F_{3,33}$ = 9.2, **p**<.0001) and Technique ($F_{2,22}$ = 6.9, **p**=.005) on effort. A one-way Anova reveals that Cheek and Oversized Temple are less tiring than Regular HWD Temple and Forehead.

While we expected this result for Forehead, which forces the user to raise her arm, the score for Regular Temple is surprising. One reason might be that the small surface requires more clutching, thus requiring the user to keep her arm suspended in air longer. The results of the Likert questionnaire reveal that users largely prefer the Oversized Temple (50% Strongly agree) and the Cheek (55% Agree). Forehead (60% Strongly disagree) and Regular temple (20% Strongly disagree and 40% disagree) were the least preferred areas by users.

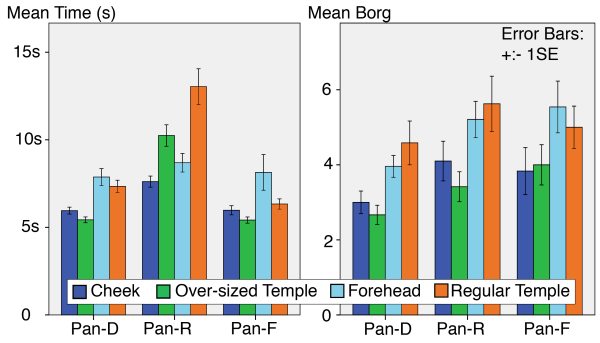


**Figure 10. Mean time in s. (left) and Mean Borg value (right) for every panning technique and interaction area.**

*Summary*

This first study on using hand-to-face gestures for panning revealed that the best facial area for input is the Cheek. The Forehead and the Regular HWD Temple not only showed worse performance, but also result in higher fatigue. Overall there was no difference between the Cheek and the Oversized temple, but both were favored over the Regular temple. The Oversized temple, however, is far larger than most HWDs, suggesting that the Cheek is a preferred interaction region.

## HAND-TO-FACE NAVIGATION STUDY 2: ZOOMING

The goal of this study is to evaluate the performance of hand-to-face gestures, for full document navigation which includes zooming and panning. From the previous study we dismiss Forehead due to its low overall results.

### Tasks and mappings

The task was the same as in the previous study, but with a much more distant target that required zooming to reach effectively. We selected three different zooming techniques, based on a combination of prior known methods and from the guessability study: Linear, Rotation and Cyclo. Linear zooming, by pinching with two fingers, is the classical touchscreen technique. Circular zooming with two fingers (using the angle of rotation) is based on the metaphor of adjusting an optical lens. Cyclo is a one finger zooming technique proposed by Malacria et al. [19] as a way to avoid clutching when zooming. It consists of doing a circular movement with one finger. The orientation of the rotation is mapped to the zoom direction (in or out). We use the same implementation as in [19].

### Participants

We recruited 15 students (2 females) from a local university with an average of 26 years. 10 of them assisted Study 1.

### Design

The experiment followed a 3×3 design with Area (Cheek, Oversized HWD temple, Regular HWD temple) and Zoom Technique (Linear, Rotation, Cyclo) as factors. We chose the Pan-D technique from the first study. We counterbalanced the Area factor, while the technique was presented by increasing level of difficulty.

### Apparatus, procedure and collected data

We used the same apparatus as in the previous experiment, adding an extra sensor and IR marker to detect thumb touch and position. We follow the same procedure and measured time, physical demand and user preference as in Study 1.

### Results

We collected 162 trials per user (3 Areas × 3 Techniques × 3 Directions × 6 Repetitions)× 12 users= 1944 trials in total.

*Time Performance*

We found a main effect of our two factors, Area ($F_{2,22} = 3.7$, **p**=.038) and Technique ($F_{2,22} = 5.5$, **p**=.009) for completion time (Figure 11). We found no interaction between Area and Technique. Post Hoc tests reveal that Cheek (mean 23.8s) is significantly faster than Regular Temple (mean 27.0s) and Oversized Temple (mean 26.7s). Concerning the Technique, 2 Finger Rotation (mean 29.7s) is significantly slower than 2 Finger Linear (mean 24.7s) and 1 Finger Cyclo (mean 23.1s). There is no significant difference between the last two techniques.

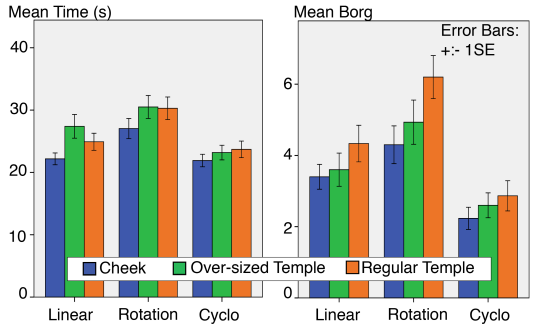


**Figure 11. Mean time in seconds (left) and mean Borg value (right) for Technique and interaction Area.**

*Fatigue and User preference*

The results of the Borg questionnaire reveal a main effect of both Area ($F_{2,22} = 8.0$, **p**=.002) and Technique ($F_{2,22} = 20.1$, **p**<.001) on fatigue. Post-hocs reveal that the Regular Temple was perceived as more fatiguing than the Cheek or the Oversized Temple (Figure 11). Concerning techniques, Cyclo was least fatiguing technique, followed by Linear and Rotation. The results of the Likert questionnaire are similar to that of study 1: overall, users preferred the Cheek (26% Strongly agree, 33% Agree) and the Oversized Temple (20% Strongly agree, 53% Agree) over the Regular Temple (26% strongly disagreed, 46% disagreed).

*Summary*

These results extend further our exploration of Study 1, providing insight into hand-to-face interaction for document navigation. The main finding is that the Cheek is more efficient than both the Oversized and Regular temples for zooming. While the Oversized temple was efficient in Study 1 for one finger panning, it becomes inefficient with a two-finger gesture. Both the classical Pinch and the single-finger Cyclo are equally efficient in our study. However, to access an intermediate zoom level Cyclo could lose precision on the Regular temple, as smaller circles cause faster zooming, which may lead to overshooting.

## SOCIAL ACCEPTANCE OF HAND-TO-FACE GESTURES

While we demonstrated that hand-to-face interaction techniques improve navigation performance, we know little on how comfortable users would feel in different social contexts. We therefore carry a controlled exploration of the social acceptance of our hand-to-face gestures.

### Questionnaire and gestures

Participants were asked to rank on a 5-point Likert scale the social acceptability of hand-to-face gestures. For every gesture and social context we asked the same question: *"Would you be willing to perform this gesture in context X?"* We grouped two social aspects into a single factor we call Social context. This factor encompasses audience (who you are with) and location (where you are), both inspired from previous studies [2, 23]. We focus on the panning and zooming gestures from the earlier studies. For panning, we include *Displacement* (Pan-D) and *Flick* (Pan-F). For zooming, we include *Linear* (L) and *Cyclo* (C).

### Participants

We recruited 12 students (5 females) from a local university with an average age of 27 years. 7 of them wear optical glasses and 8 of them had tested a head-worn device. None of them participated in our previous studies. They were rewarded with a drink.

### Design

We used 2×8×6 within-participant design with Device (Face or HWD), Gesture (Pan-D and Pan-F panning, L and C zooming) and Social context (Alone, Family or friend, Stranger, Home, Public space and Workspace) as factors.

### Procedure

The study was performed in the presence of the interviewer. Participants watched a video of an actor performing panning and zooming gestures in front of a wall and then performed themselves the same gestures 3 times. The order of the videos was counterbalanced between participants. After completion of all the gestures, participants completed a questionnaire containing one question for each gesture and social context. This type of exploration based on video watching has already been used in previous studies [2]. Although it misses the ecological validity of an experiment in a real environment, it allows for a first exploration of a novel technique. For every participant we collected 48 responses to the Likert scale questions (2 × 8 gestures × 6 social contexts), written comments and oral feedback.

### Results

We collected 576 answers (48 answers × 12 subjects) for the 5-points Likert scale questionnaire (5= full agree).

*Comparison of face and HWD acceptability*

We used a Wilcoxon signed-ranks test to compare our main factors. We found do difference in social acceptability between Face and HWD ($Z=-1.817$, $p=.069$), but with a trend showing better acceptance for interaction on the HWD. The acceptance rate for both face and HWD gestures in any social context is above 50% (Figure 12). Results were rather homogeneous on the HWD, with a constant 10%-12% of disagreement for all social contexts except in front of strangers, where this value is 18%. We found more differences on the Face, with no disagreement when at Home or Alone, but with 31% disagreement in Public places and 25% in front of strangers. Comments from participants also show that most of them don't mind using the face: *"I don't think it would disturb me to do the gesture either on the skin or on the temple."* [P6]. One female participant indicated the problem of dirty hands on the face: "*the face can be affected when perspiring*" [P7].

*Hand-to-face gestures acceptability*

A Friedman test reveals a significant difference in social acceptability among techniques ($p<.001$). We used a Wilcoxon signed-rank test for our analyses. We found a significant difference between the panning techniques ($Z=-2.06$, $p=.039$), Pan-D being better perceived than Flick. We also found a difference between the zooming techniques ($Z=-6.2$, $p<.001$), L being better perceived than C. Participants commented that Cyclo might be perceived as insulting, as it could signal that "*you are crazy*" [P4] in many cultures. This gesture seemed also more visible: "*I feel all the gestures are quite subtle except Cyclo which might attract attention*" [P6].

## DISCUSSION

### Extending HWD through Hand-To-Face Interaction

Interacting with HWDs via the device temple presents a number of limitations, such as the small surface, movement due to contact and user fatigue. While Face interaction is not meant to replace HWD interaction, particularly for rapid selection, it is a promising alternative for extended tasks such as navigation. In our experiments we discovered that face interaction offers the benefit of faster panning + zooming as well as lower fatigue than the HWD.

### Hand-To-Face gestures acceptability

While our social acceptability study is a laboratory exploration that should be expanded to the real environment, it indicates that hand-to-face interaction could be acceptable in different social contexts. Users find calm gestures such as displacement or pinching to be more socially acceptable than brisk ones such as Flick. Unexpectedly, the Cyclo gesture which ranked among the

most efficient from study 2 was seen as not acceptable. Users also identified some potential resistance to using Hand-to-Face input, due to practicalities such as facial cosmetics. One alternative to explore would be mid-air gestures near, but not touching the face. However, mid-air gestures need to be carefully designed to include proper delimiters, to not appear socially unacceptable, and to be evaluated against on-face input as a touch-less solution for addressing the issue of facial cosmetics.

### Hand-To-Face detection

We explored hand-to-face gestures for HWDs without emphasizing the technology that would ultimately support this style of input. Several options exist, such as a camera mounted on the HWD [15], body-implanted sensors [13] or instruments worn on the finger [34]. Ideally such instruments should consider contactless hand-to-face input, as a means to avoid the limitations discussed above.

### Casual and explicit hand-to-face interaction

Hand-to-face gesture recognition will need to consider the Midas-touch problem [12], i.e. how to differentiate casual and explicit hand-to-face gestures. Two obvious solutions include touching the HWD to initiate the face detection or using a voice command. Another solution is to use gestures that are very different from casual ones. While casual gestures are rather static and use the whole hand [18], gestures such as pinching could be easier to differentiate. Further investigation is needed to explore the possibilities.

### Other applications for hand-to-face interaction

While hand-To-face interaction appears to be particularly well suited for HWDs, other contexts may benefit from hand-to-face interaction. One of the main benefits of hand-to-face interaction is to allow interaction with one hand when no input device is available. It could thus be used in mobile situations, such as on a bike to control a music player or when the other hand is busy holding a handle bar on a bus. Further work is needed to assess the value of hand-to-face interaction in eyes-free situations.

### Lessons learned

This first exploration allows us to sum up a number of prescriptive design recommendations. Gestures should be performed on the cheek or on the lower areas of the face (chin, jaw), which are less tiresome than the Forehead. While the cheek has demonstrated value for navigating a document, other areas of the face can be used to enhance the symbolism and memorization of certain tasks. For instance, touching the chin can be used to activate a command related to speech, such as making a call. Hand-to-face gestures should be calm, since obvious or vigorous gestures can be less socially acceptable.

### Gesture alternatives

Prior work on user-generated gestures, indicated that designers came up with only 60% of the gestures produced by end-users [33]. This gives some evidence that users may, in some instances, generate a larger set of alternatives than a few designers. Some gestures may have been missed though, such as using pressure for zooming. Our final gestures resemble known multi-touch gestures as we had to select a set that was compatible with our baseline, the HWD temple: given its limited surface area, our choice of gestures was limited. In future, we intend on exploring other interactions without the baseline restrictions.

### Limitations and future work

Our exploration was limited by the apparatus used in our experiments, by the participants profile and by the context of our social acceptability study. The overall apparatus, including the large HWD and the glove, was bulky and heavy, although it allowed us to demonstrate the benefits of hand-to-face interaction. While these benefits should be further noticed with lighter form factors (due to limited temple area), further evaluation is needed. Our comparison focused mainly on the impact of the temple's size. However other aspects such as the HWD's grip on the user's head need to also be considered when characterizing interaction with a HWD. In particular, the results of our guessability study apply best to HWDs with a similar form factor as our mock-up prototype. In future, we will replace the heavy HWD with an improved, lighter version. We also plan to integrate the sensing technology on the HWD itself, for instance using a camera fixed on its frame (similar to [15]).

Our participants were mainly local students. The findings of our work should be verified with a more diverse population, particularly for better understanding guessability and social acceptability issues. Our social acceptability study was based on video browsing gestures. Future work should test the acceptability of such interactions in real settings, such as in [22]. Factors such as gender and cultural ethnicity also need to be considered for improved Hand-to-Face input. For instance we can compare users' perceived acceptability between Asian and Western populations as in [22].

## CONCLUSION

We presented an exploration of hand-to-face gestures, a novel type of on-body interaction especially well-suited for HWDs. We first described the results of a guessability study that shows that for navigation tasks such as panning and zooming participants prefer using hand-to-face gestures rather than gestures on the HWD. We calculated a similarity score among these gestures that demonstrates that participants converged to similar hand-to-face gestures for panning and zooming. In a first study we explore different areas and techniques for panning. We found that the cheek is the most promising area on the face, being larger, more efficient, less tiring and preferred to a regular temple on the HWD. In a second study we investigate different techniques for zooming and show that using the cheek is more efficient than using the temple of the HWD. Finally, we enquire into the social acceptability of these gestures and show that gestures on the face could be as acceptable as on the HWD.

## REFERENCES

1. Alexander, J., Han, T., Judd, W., Irani, P. Subramanian, S. 2012. Putting your best foot forward: investigating

real-world mappings for foot-based gestures. In *Proc. of CHI '12*. ACM, 1229-1238.

2. Bailly, G., Müller, J., Rohs, M., Wigdor, D., Kratz, S. 2012. ShoeSense: a new perspective on gestural interaction and wearable applications. In *Proc. of CHI '12*. ACM, 1239-1248.
3. Borg, G. Borg's Perceived Exertion and Pain Scales. Human Kinetics (1998), viii 104pp.
4. Epson Moverio, http://www.epson.com/moverio
5. GlobalWebIndex, https://www.globalwebindex.net/Top-global-smartphone-apps
6. Google Glass, Project Glass One Day video, http://www.google.com/glass
7. Gustafson, S.,Rabe, B., and Baudisch, P.. 2013. Understanding palm-based imaginary interfaces. In *Proc. of CHI '13*. ACM, 889-898.
8. Harrison, C., Tan, D., and Morris, D. 2010. Skinput: appropriating the body as an input surface. In *Proc. of CHI '10*. ACM, 453-462.
9. Harrison, C., Benko, H. and Wilson, A. 2011. OmniTouch: wearable multitouch interaction everywhere. In *Proc of UIST'11*. ACM, 441-450.
10. Hidayat, A. Flick list with its momentum scrolling and deceleration. http://ariya.ofilabs.com/
11. Hincapie-Ramos, J-D,. Guo, X., Moghadasian, P. and Irani, P. 2014. Consumed Endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proc. of CHI'14*. ACM, to appear.
12. Hinckley, K., Wigdor, D., Input Technologies and Techniques. Chap. 9 in The HCI Handbook, 3rd Edition, Taylor & Francis.
13. Holz, C. Grossman, T., Frizmaurice, G and Agur, A. 2012. Implanted user interfaces. In *Proc. of CHI '12*. ACM, 503-512.
14. Kim, D. Hilliges, O., Izadi, S., Butler, A., Chen, J., Oikonomidis, I. and Olivier, P.. 2012. Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *Proc. of UIST '12*. ACM, 167-176.
15. Kimura, S., Fukuomoto, M., and Horikoshi, T. 2013. Eyeglass-based hands-free videophone. In *Proc. of ISWC '13*. ACM, 117-124.
16. Lissermann, R.Huber, J., Hadjakos, A. and Mühlhäuser, M. 2013. EarPut: augmenting behind-the-ear devices for ear-based interaction. In *CHI '13EA*. ACM, 1323-28.
17. Lucero, A. Lyons, K., Vetek, A., Järvenpää, T., White, S. and Salmimaa, M. 2013. Exploring the interaction design space for interactive glasses. In *CHI '13 EA*. ACM, 1341-1346.
18. Mahmoud, M. and Robinson, P. 2011. Interpreting Hand-Over-Face Gestures. In *Proc. of Affective Computing*, Springer, 248-255.
19. Malacria, S., Lecolinet, E. and Guiard, Y. 2010. Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces: the cyclostar approach. In *Proc. of CHI '10*. ACM, 2615-2624.
20. Nicas, M. and Best, D. 2008. A Study Quantifying the Hand-to-Face Contact Rate and Its Potential Application to Predicting Respiratory Tract Infection. In *J. Occup. Env. Hyg.*, 347-352.
21. Perrault, S., Lecolinet, E., Eagan, J., Guiard, Y. 2013. Watchit: simple gestures and eyes-free interaction for wristwatches and bracelets. In *Proc. of CHI'13*. ACM, 1451-1460.
22. Profita, H., Clawson, J., Gilliland, S., Zeagler, C., Starner, T., Budd, J. and Do, E. 2013. Don't mind me touching my wrist: a case study of interacting with on-body technology in public. In *Proc. of ISWC '13*. ACM, 89-96.
23. Rico, J. and Brewster, S. 2010. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proc. of CHI '10*. ACM, 887-896.
24. Ruiz, J, Li, Y., Lank, E. 2011. User-defined motion gestures for mobile interaction. In *Proc. of CHI '11*. ACM, 197-206.
25. Saponas, T., Kelly, D., Parviz, B. and Tan, D. 2009. Optically sensing tongue gestures for computer input. In *Proc.of UIST '09*. ACM, 177-180.
26. Serrano, M., Lecolinet, E., and Guiard, Y. 2013. Bezel-Tap gestures: quick activation of commands from sleep mode on tablets. In *Proc. of CHI'13*. ACM, 3027-3036.
27. Seyed, T., Burns, C., Costa, M, Maurer, F. and Tang, A. 2012. Eliciting usable gestures for multi-display environments. In *Proc. of ITS '12*. ACM, 41-50.
28. Špakov, O. and Majaranta, P. 2012. Enhanced gaze interaction using simple head gestures. In *Proc. of UbiComp '12*. ACM, 705-710.
29. Tamaki, E. Miyak, T. and Rekimoto, J. 2010. BrainyHand: a wearable computing device without HMD and its interaction techniques. In *Proc. of AVI '10*. ACM, 387-388.
30. Telepathy, http://tele-pathy.org
31. Vuzix M100, http://www.vuzix.com
32. Wagner, J. Nancel, M., Gustafson, S., Huot, S. and Mackay, W. 2013. Body-centric design space for multi-surface interaction. In *Proc. of CHI '13*. ACM, 1299-1308.
33. Wobbrock, J., Ringel Morris, M. and Wilson, A. 2009. User-defined gestures for surface computing. In *Proc. of CHI '09*. ACM, 1083-1092.
34. Yang, X-D. Grossman, T., Wigdor, D. and Fitzmaurice, G. 2012. Magic finger: always-available input through finger instrumentation. In *Proc. of UIST '12*. ACM, 147-156.