Designing Socially Acceptable Hand-to-Face Input

DoYoung Lee, Youryang Lee, Yonghwan Shin, and Ian Oakley
Human Factors Engineering, UNIST, Ulsan, Republic of Korea
dylee.issac@gmail.com, {yrlee, yonghwanshin}@unist.ac.kr, ian.r.oakley@gmail.com

ABSTRACT
Wearable head-mounted displays combine rich graphical output with an impoverished input space. Hand-to-face gestures have been proposed as a way to add input expressivity while keeping control movements unobtrusive. To better understand how to design such techniques, we describe an elicitation study conducted in a busy public space in which pairs of users were asked to generate unobtrusive, socially acceptable hand-to-face input actions. Based on the results, we describe five design strategies: miniaturizing, obfuscating, screening, camouflaging and re-purposing. We instantiate these strategies in two hand-to-face input prototypes, one based on touches to the ear and the other based on touches of the thumbnail to the chin or cheek. Performance assessments characterize time and error rates with these devices. The paper closes with a validation study in which pairs of users experience the prototypes in a public setting and we gather data on the social acceptability of the designs and reflect on the effectiveness of the different strategies.

Author Keywords
Hand-to-Face Input; Social Acceptability; User Elicitation; Augmented Reality; Head Mounted Display

INTRODUCTION

Augmented Reality (AR) Head-Mounted Displays (HMDs) are an emerging consumer technology that promise to impact activities as diverse and fundamental as education [27], accessibility [14], health care [37] and entertainment [34]. Understandably, AR has long attracted attention in the Human-Computer Interaction (HCI) research community [3]. However, while aspects such as tracking fidelity, display quality and computing power have advanced considerably to produce today's high-end products, input and interaction technologies are less mature. Current commercial systems feature on-headset touch surfaces (e.g., Google Glass) or hand controllers in the form of touchpads (e.g., the Epson BT-300) or hand-held clickers (e.g., Microsoft HoloLens) as key interaction channels. While these systems can be effective, they offer limited input areas and, in the case of hand-held controllers, are cumbersome additional devices that disrupt or preclude system use during mundane, everyday tasks and activities in which the hands are busy.

Recognizing the need for input systems for AR glasses that leave the hands unencumbered, a considerable body of research has explored topics ranging from wearable peripherals, such as belts [8] or rings [9], to in-air gestural input [13] and on-body touches [35]. One focus for work in this latter area has been on using the face as a site for input – what Serrano *et al.* [31] term *hand-to-face* input. The face is appealing as it is easy to access with a touch, typically unobstructed by garments and proximate to smart glass hardware. Facial touching is also a common human behavior [24]. Prior work has shown that input on the face can be useful in tasks such as navigation, video browsing [31] or basic selection and pointing [32] through schemes such as swiping and tapping on the cheek [31], thumbing the nose [19] and stroking the hair [6].

While this work effectively demonstrates the viability and diversity of hand-to-face input, it is also fragmented, piecemeal and technologically opportunistic. By this we mean that proposals typically target highly specific body sites such as the ear [17], nose [19], cheek [43] or hair [6] with the goal of exploring interactions that can be effectively performed by users and detected by a predetermined sensor setup, such as electrooculography (EOG) glasses [19], optical range finder arrays [43] or capacitive braids [6]. We argue there is a need to improve our understanding of how users conceive of touches to the face as an input modality [31] to better inform future design and development efforts. Specifically, we argue that a key omission in our current understanding relates to the social acceptability [29] of facial touches - how comfortable users feel performing or observing this type of input in real life situations. This issue is particularly important as the face is an exposed, publicly visible body region and AR systems are ultimately intended and expected to be used in everyday settings and spaces, situations in which many forms of publicly observable input may be considered socially unacceptable [13].

We studied this issue in a multi-stage research process. First, we conducted an elicitation study [41] of input via facial touches with pairs of users in a public setting – a coffee shop. Users created and rated interface proposals according to how comfortable they felt performing them in various settings [1]. We contribute both this novel combination of elicitation and social acceptability methods and the study results in the form of strategies for designing socially acceptable hand-to-face input techniques for AR/wearables. Building on these strategies, we then created, implemented and evaluated two input systems, one based on camera tracked touches to the ear, the other on a touch-sensitive thumb nail. These activities con-

UIST 2018, October 14–17, 2018, Berlin, Germany.

ACM ISBN 978-1-4503-5948-1/18/10 ...$15.00.
http://dx.doi.org/10.1145/3242587.3242642

tribute a novel sensing setup (camera tracking of touches to the ear) and input approach (nail to face touches) as well as empirical characterizations of user performance and recommendations for how these systems be configured to support effective, expressive input. We close the paper with a study in which participants used our techniques in a public setting and contribute a qualitative assessment of their social acceptability that serves to validate the techniques with respect their original design intentions. In this way, we showcase the value of our design strategies for creating socially acceptable hand-to-face input techniques for AR/wearables.

RELATED WORK

There is a large and rapidly growing literature dealing with on-body input. One key focus is on developing sensing solutions based on, for example, signals that propagate through the body [11], camera-based tracking [4], or thin sensing films [39]. While most of this work focuses on point touches, researchers have also emphasized that much more can be done with the skin. It is highly flexible and readily deforms, providing additional channels for input [26]. For example, Weigel *et al.* [38] applied elicitation methods to understand the potential of skin deformation for interaction design, highlighting its emotionally rich and evocative qualities.

While the majority of this work focuses on the hand or forearm, touches to the face are a common behavior that interaction designers can also leverage. Basic studies of face touching behavior indicate it occurs frequently – at rates of 15.7/hour for the mouth, eyes and nose [24] through 24/hour [18], 40/hour [7] and up to 54.3/hour for the whole face [12]. In early work to explore the value of these touches for device input, Serrano *et al.* [31] conducted an elicitation study and concluded that fairly standard finger strokes (swipes, two-finger pinches and circles) on the cheek were an optimal design candidate. Subsequent studies characterized empirical input performance (using a high end optical motion capture system) and a lab-based assessment of social acceptability indicated participants felt that smaller and simpler gestures such as swipes were unlikely to attract undue attention.

Recent work has focused on taking hand-to-face input out of the lab by broadening the design space and constructing viable sensing systems. This later task is challenging and the expressivity of current systems is low. For example, Lisserman *et al.* [20] created an array of capacitive sensors that fit behind the ear; they reported that two touches to the ear can be detected accurately (99%) but performance with three or more touches drops steeply (to 86.6% or lower). Kikuchi *et al.* [17] expand on these ideas with a system that uses in-ear optical sensors to detect five ear deformations – four directional pulls and a press – with an accuracy of up to 89.56%. Yamashita *et al.* [43] describe a broadly similar system that deploys an array of optical sensors embedded in glasses, and focused on the skin of the face, to track five pushes to the cheek with an accuracy of 89.8%. In closely related work, Lee *et al.* [19] use off-the-shelf EOG glasses to detect five actions (rubs and left/right flicks and pushes) on the nose with an accuracy of up to 96%. Finally, Dierk *et al.* [6] present a proof of concept system featuring actuated hair braids that use swept frequency capacitive sensing to detect touches along their length.

While much of the practical work in these projects involves building effective input technologies, a key goal underlying their explorations of hand-to-face input is the idea that it is an appropriate way to interact in public settings. For example, Lee *et al.*'s [19] primary motivation for their system is to create discreet input primitives, while Dierk *et al.* [6] stress the public and social aspects of hair in the design guidelines that informed their prototype and Kikuchi *et al.* [17] are inspired by the idea that the ear can be touched "naturally without worrying about provoking stares". While the inherent unobtrusiveness, subtlety or social acceptability of hand-to-face input is an appealing idea that is well grounded in the literature capturing the regularity of facial touches, we note is a largely unexamined assertion – it has tended to be claimed rather than assessed. In other words, despite a growing research interest in developing hand-to-face input systems, there is a lack of design knowledge about the types and forms of hand-to-face touch that are appropriate for public settings.

Subtlety, Unobtrusiveness and Social Acceptability

Although they have seen scant attention in studies of hand-to-face input, issues of subtlety, unobtrusiveness and social acceptability have attracted research attention in closely related areas. Early work emphasizing the subtlety [5] and unobtrusiveness [28] of input actions focused on wearable technology and argued these qualities are requirements to achieve social acceptability. The subtlety of an interface has also been operationalized as the frequency with which it can be used without an observer noticing [2]. Building on these ideas, social acceptability has been defined as whether an input action is deemed appropriate by both the user issuing it, and observers watching it, in the context in which it occurs [21]. It has been applied to assess the viability of general body gestures (e.g., foot tapping or head nodding) and mid-air gestures of [29], or around [1], mobile devices. This work typically shows a predetermined set of gestures to participants via video, or requires them to enact such a set. Participants then rate where (e.g., home, restaurant, workplace) and in front of whom (e.g., alone, with family, friends, strangers) they would either be willing to (or feel comfortable performing) each gesture. This approach has generated a range of useful outcomes. Rico *et al.* [29], for example, highlight the importance of subtle, comprehensible, and familiar movements in maximizing social acceptability, while Ahlström *et al.* [1] provide concrete recommendations for creating socially acceptable in-air gestures based on pragmatic qualities such as gesture size, location, and duration and Montero *et al.* [21] provide a useful classification of gesture social acceptability based on the whether the input actions and/or resultant outcomes are observable.

HAND-TO-FACE ELICITATION STUDY

Elicitation studies involve participants creating input actions for a given set of tasks [41]. The resultant actions are analyzed for factors such as their agreement across participants [33] and categorized according to salient design properties, such as their form or complexity [30]. This study sought to understand how users conceive of hand-to-face input in a social setting. It moves beyond prior work [31] by focusing on social acceptance and unobtrusive or subtle actions during the input creation phase and by including an evaluation of these qualities by both participants and observers. In this way, it seeks to generate design knowledge that can support creation of novel socially acceptable input techniques for AR/wearables.

Experimental Design, Tasks and Setting

Study tasks were adapted from prior work [41, 31] and are shown in Table 1; each participant was asked to generate hand-to-face input actions for each task. Following Morris [22], we deployed three methods to improve study outcomes: production, priming and partnership. Production relates to generating multiple proposals; participant's generated two and selected a favorite. Priming involves providing illustrations and examples to help participants move beyond existing designs from other contexts. We achieved this by showing participants a 90 second video depicting a wide variety of hand-to-face input actions, which they were encouraged to try out, and giving them a demo of the Microsoft HoloLens. Finally, partnership relates to performing elicitation activities in groups. We achieved this by recruiting pairs of participants (all strangers) and having them generate interface proposals alternately – this enabled them to build and reflect on their partner's ideas.

To focus the study on social acceptability, we further adapted typical elicitation methods. To improve the ecological validity of the proposed actions, the study was conducted in a busy public place – a coffee shop. Participants were also instructed to generate unobtrusive or subtle actions, suitable for use in the public setting of the study, and both they and their partners rated all favored input proposals for social acceptability and how obvious the input action was. This provided both author and observer perspectives on social acceptability [29]. To assess social acceptability, we used Ahlström *et al.*'s [1] questionnaire; this asks what situations and individuals a participant would feel comfortable performing an action in. For obviousness (or unobtrusiveness), we adapted the social acceptability questionnaire to ask in which locations an action would be obvious. We also included a seven-point Likert scale rating actions from "not obvious" to "very obvious". In total, participants answered four questions per favorite input action generated by both themselves and their partner.

Participants and Procedure

Twenty participants (10 males, mean age 22.6, one left-handed) completed this study in ten pairs. All were either students or recent graduates at UNIST and they were compensated with approximately $15. Participants self-rated as highly familiar with smartphones (4.5/5) and computers (4.5/5), but had no or limited experience of HMDs and AR. Each study session began with an introductory video showing example actions and a demo of the HoloLens. Participants then began generating and rating input actions, alternating making the initial proposal with their partner. Half the participant pairs generated proposals for tasks in the order shown in Table 1 and half in the inverse order. While generating proposals, they were encouraged to think aloud in order to expose their design intentions to both their partner and the experimenters. After both participants in a pair had generated two proposals and selected a favorite, they rated these for social acceptability and how obvious they were. The experiment took approximately two hours for each pair and, in total, 320 favorite gestures were produced: 16 tasks by 20 participants.

Table 1. Task list used in elicitation study

Task Type	Tasks		
System	Open Delete Copy	Close Accept Paste	Select Decline Take Photo
Navigation	Next Zoom in Rotate anti-clockwise	Previous Zoom out	Pan (any direction) Rotate clockwise

Results and Data Processing

Favored actions were evenly split between first (57%) and second (43%) proposals, highlighting the value of Morris *et al.*'s [22] production recommendation. We first calculated agreement over the proposals [41] attaining an average score of 0.082 which indicates relatively low agreement throughout the study. Agreement scores peaked at 0.11 for the select task; five participants chose the action of tapping the cheek. As the goal of this study was to generate diverse proposals relating to socially acceptable input actions, and participants' task instructions reflected this, we do not view the low agreement scores as problematic; the diversity they hint at is more appropriate for our goals. Based on our goals and this outcome, we opted not to calculate a consensus set of input actions.

We next classified actions according by the *face site* and *hand action* used (Table 2). As in prior work [31], the cheek was popular (27.2% in the current study vs 34% in prior work), but our data saw more emphasis on the chin (18.2% vs 7%) and ear (10.3% vs 7%) over areas such as the forehead (5.6% vs 16%). This again likely reflects the differing experimental instructions – in the current study participants tended to avoid prominent features such as the forehead in order to generate socially acceptable or non-obvious/unobtrusive actions. Hand actions tended to be swipe-like strokes (27.9%), or various forms of tap (e.g., tap, push, long-tap: 29.9%) over more unusual types of input. A long tail of alternatives for both classifications doubtless contributed to the low agreement scores.

To aid in our interpretation of this data, we derived numerical *acceptability scores* from the questionnaire results from all participants and input proposals and used these figures to calculate means for each face site and hand action category. Acceptability scores were generated as follows. For the three nominal questions, participants checked between zero and seven/eight options to represent either the situations or individuals/groups they would feel comfortable performing an input action in or in front of [1] and the situations in which an input action would be obvious. For each question, we calculated the percentage (0-1) categories selected. The Likert scale captured how obvious input actions were (0-7 scale). Based on the idea that selecting more locations/situations indicates increased social acceptability, the acceptability score was the mean of the percentages from the nominal questions and the normalized (0-1) score from the Likert scale. The results are shown in

Table 2. In terms of face region, the highest scores are on the ear, neck and temple, which suggests that input areas away from the center of the face may be appropriate for hand-to-face designs. The lowest scores are reserved for areas such as the hair, whole face or nose, regions where input actions were either large or front-and-center. These trends were also evident in the scores for hand actions – tap, a small discreet movement, rated best, while spread, a large action involving moving all five fingers out from a central pinch, scored poorly. Similarly, actions such as fold, universally applied to the out-of-the-way ear, scored highly. While we believe these ratings are useful, one caveat to their interpretation is that they are based on varying numbers of example proposals. For example, the rating for swipe is the aggregate of many (27.9%) proposals, while the score for fold is derived from a handful (1.7%). Variations in the number of input proposals in any given category is an inevitable outcome from an elicitation study.

Generation Strategies

Building on these analyses and classifications, we sought to understand participants' strategies for creating socially acceptable or unobtrusive inputs by constructing a taxonomy, a common outcome from elicitation studies [30, 41], of their approaches. To achieve this goal, two experimenters reviewed the proposed actions, think aloud notes, summary statistics and questionnaire data independently, then created initial categorizations and discussed the outcomes until they reached consensus. Ultimately, of the 320 proposals, 83 were unclassified and 237 were assessed as exemplifying one or more of the following five strategies (bracketed figures show count, mean acceptability score): miniaturizing (51, 0.72); obfuscating (24, 0.76); screening (27, 0.71); camouflaging (185, 0.71) and; re–purposing (39, 0.69). The 83 unclassified gestures achieved a mean acceptability score of 0.6 and 58 actions fell into two categories, 11 into three and three into four categories.

The first three strategies involve hiding input actions. *Miniaturizing* is simple. It relates to keeping movements as small as possible and is closely related to previously observed strategies to achieve social acceptability in gesture input [29]. *Obfuscating* represents the use of a face region that is naturally hidden from observers in front of a user, such as the back of the neck or, when the head is turned, the ear. *Screening* involves the use of the hand to hide the input action, such as movements of the thumb on the face while it is obscured behind the fingers. The remaining two strategies seek to avoid arousing attention, even if actions are observed. *Camouflaging* refers to creating input actions based on unconscious or non-communicative facial touches such as scratching the face or running a hand through the hair. Finally, *re–purposing* entails using clear and explicit intentional actions, such as nodding the head (when touching the chin) or purposefully adjusting the hair. Figure 1 shows representative examples of these five strategies.

Table 2. Distribution and acceptability scores for face region and hand action in the elicitation study. Unless otherwise specified (e.g. palm, back), all hand actions involved the finger. The 23 hand actions that were proposed in less than 1.5% of input proposals are not shown.

Face Region	*Selection Freq.*	*Acceptability Score (0-1)*	*Hand Action*	*Selection Freq.*	*Acceptability Score (0-1)*
Cheek	27.2%	0.75	Swipe	27.9%	0.76
Chin	18.2%	0.75	Tap	16.6%	0.81
Ear	10.3%	0.79	Push	7.3%	0.67
Hair	7.0%	0.67	Long-Tap	6.0%	0.77
Lip	5.6%	0.71	Grab	4.3%	0.75
Forehead	5.6%	0.70	Palm-Swipe	4.0%	0.76
Neck	5.3%	0.80	Pinch	3.7%	0.79
Temple	4.3%	0.79	Flick	3.3%	0.76
Nose	4.3%	0.70	Palm-Place	3.0%	0.76
Cheekbone	3.6%	0.78	Pull	3.0%	0.75
Eye	3.6%	0.72	Palm-Push	2.0%	0.78
Eyebrow	3.3%	0.76	Spread	2.0%	0.67
Whole Face	1.7%	0.70	Fold	1.7%	0.80
			Back-Swipe	1.7%	0.73
			Twist	1.7%	0.70

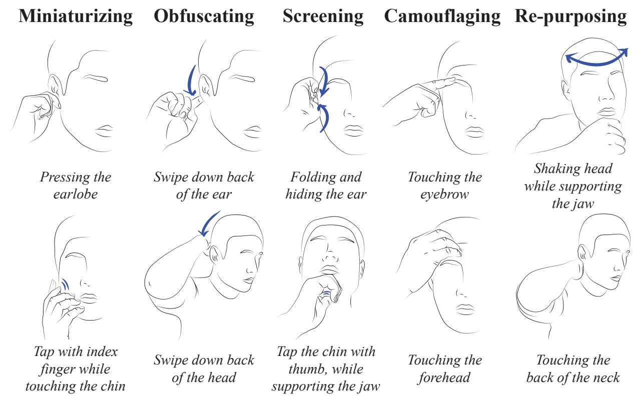

Figure 1. Examples of the five design strategies. *Miniaturizing* illustrates small finger movements, such as taps on or presses to the skin. *Obfuscating* shows input actions intended to be hidden to the side of or behind the head. *Screening* involves concealing input movements within or under the hand. *Camouflaging* masks input actions in seemingly unconscious movements, such as rubbing the eyebrow or scratching the forehead. *Re-purposing* relates to co-opting clear and explicit gestures, such as nodding the head or massaging the neck, as control inputs.

CASE STUDIES

To explore the value of the data and design strategies captured in, and derived from, the elicitation study, we developed two prototype hand-to-face input systems. These designs directly reflect outcomes from the elicitation study: they use the most commonly proposed face sites and finger actions; they instantiate different combinations of the design strategies and; each prototype was adapted from specific user proposals (see Figure 1 for examples). For each prototype, we describe the designs, implementations and empirical studies that assess if they are effective at supporting hand-to-face input tasks.

Case Study 1 - EarTouch

The first case study is a system for input on the skin of the ear; it builds on prior work that has explored input on this body site [20, 17]. This site was selected as the ear was the third most popularly selected face site in the elicitation study and it embodies several of the design strategies. Specifically, it supports *obfuscating* input, as the ear can be hidden from an observer simply by turning the head, *miniaturizing*, as the available space for input on the ear is small, and *camouflaging* in that input actions could be disguised as common actions such as scratching. Additionally, prior work on ear based input has achieved quite limited expressivity, reporting reliable recognition of between two [20] and five [17] input actions. Accordingly, we sought to boost this performance.

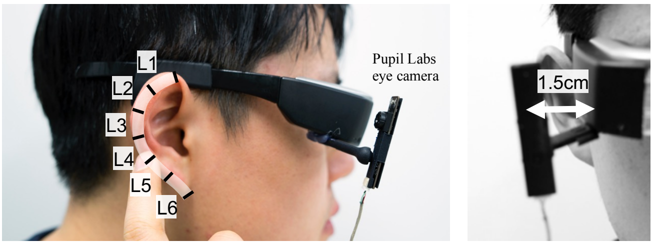

Figure 2. EarTouch device prototype. Left: targets used in the study shown mapped on the ear. Right: tracking camera mounted to the HMD.

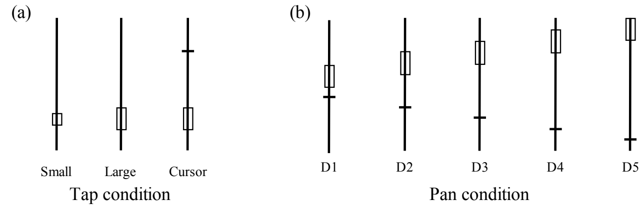

Figure 3. Interface for EarTouch study showing (a) tapping task interface for size and feedback variables and (b) panning tasks interface for shortest (D1) to longest (D5) distances in the upwards direction.

To achieve this our EarTouch implementation used a high resolution, low latency camera tracking system. While this approach borrows from wearable input devices for the hand [4], we know of no prior systems that have used cameras to track touches to the ear. We argue a camera based approach will increase sensing resolution while reducing weight and hardware complexity compared to existing capacitative sensing implementations [20] and that it can support detection of a broader range of input actions (e.g. tap, dwell, swipe) than prior work on in-ear sensing [17], which is basically restricted to detecting large ear deformations due to actions such as bending or squeezing. Our implementation leveraged the fact that current AR HMDs protrude from the side of the head at the temple (e.g. HoloLens, 2.5cm) and involved mounting a backwards facing Pupil Labs eye camera [16] on a Epson-BT200 AR HMD to capture a clear view of the ear - see Figure 2. We expect a camera to track the ear could be mounted within the housing of many current HMDs. Although originally intended for eye-tracking, the Pupil camera's ball-and-socket joint and adjustable focus lens allowed us to readily track the ear. We configured the camera to capture 480x270 pixel images at 30Hz and connected it to a PC to perform image processing and extract ear touches. The PC transmitted the resultant data to the Epson glasses in real time via OSC/UDP over WiFi; latencies were less than 7ms, creating a smooth user experience.

The image processing system to extract ear touches was straightforward. After participants donned the glasses, we manually adjusted the camera to capture the region around the entire ear. We then extracted the ear's rough outline by segmenting the video feed by optical flow regions during head movement or when objects were moved directly behind the head. In such situations, the view of the ear is static, while the background moves rapidly. We use this initial ear region to sample the ear's color hue and refine the area based on color segmentation. Using these techniques, we achieved an easy to acquire, reliable and stable model of the shape of the ear. Additionally, we were able to re-use the hue data from the ear to capture the location of touching fingers as they came into contact with the ear; ear and finger skin were relatively similar in hue. Based on this approach, we defined the touch contact location as the normalized point at which the largest finger (or skin hue colored region) intersected the ear outline. To smooth irregularities in this data, we applied a linear easing function prior to sending it to the Epson glasses. We also discarded the first and last 100ms of finger contact data in order to focus on the stable central period of each touch [36].

Performance Study

To evaluate performance of this system we used two basic input tasks: tap (selecting a target) and pan (moving a cursor from a start to an end location). For the tap task we studied three variables: *tap-technique* (three levels); target *size* (two levels) and; target *location* (six levels). The three tap-techniques were adapted from work on HMD touch input [10]. They were *landOn*, which triggered a selection event on an initial touch, *liftOff*, which triggered selections on removing a finger and *dwell*, which triggered selections after a participant remained on a target for 400ms (a typical value for touch input [23]). Levels for the target size and location were selected (via pilot tests) to be challenging but achievable. The six target locations were equidistantly spread along the whole ear while the two target sizes were *large*, set at 1/6 of the ear size, and *small*, set as 1/12 of the ear size. The interface for this task is shown in Figure 3.a. A line depicts the ear, with the current target area highlighted. In the liftOff and dwell conditions a cursor showed the location of a touch in this space.

The panning task followed a similar design: three *pan-techniques* by five pan *lengths* by two pan *directions*. The pan-techniques were derived from prior work on hand-to-face input [31, 20]. They were: *drag*, a zero-order control method where the on-screen cursor position changed directly with finger movement on the ear; *joystick*, a first-order control method where finger displacement controlled the rate of cursor movement and; *toggle*, a system that divided the ear into three equal regions, each with a different function. Touching the top third of the ear moved the cursor upwards at a fixed speed, the bottom third moved it downwards and a location was selected by a touch to the center. In drag and joystick, a location was selected by releasing the touch. The position/speed mapping for the joystick (twice displacement/second) and fixed speed used in toggle (50% of ear length/second) were set via pilot testing. This task used the six large targets from the tap task, leading to a total of five possible distances between targets in two possible directions. The interface is shown in Figure 3.b. Participants in all conditions were required to touch their ear and then adjust their finger position to move the cursor from its initial location to the displayed target location.

Participants completed tap then pan tasks. Within each task, the study used a fully balanced repeated measures design for the technique variables and a partially balanced design for the binary variables of size and direction. For each combination, all target-position or pan-length trials were shown in a random order to form a single block of trials. For both tasks, each block

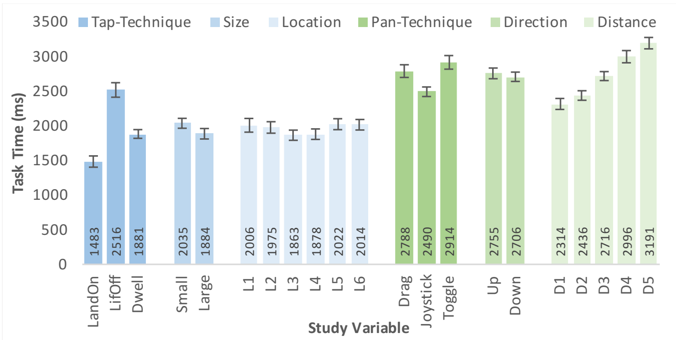

Figure 4. Task times in EarTouch tap & pan tasks. Bars show Std Err.

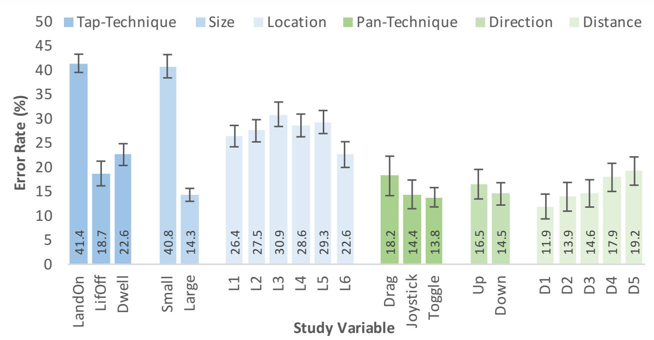

Figure 5. Error rates in EarTouch tap & pan tasks. Bars show Std Err.

was repeated four times, with the first presentation discarded as practice. Failed trials were repeated. Measures were task time (from presentation of the instructions until selection), error rate and selected point. To minimize fatigue, participants took a short rest break after each technique condition.

A total of 18 participants were recruited (14 males, mean age 24.7, one left-handed, all UNIST students) and compensated with approximately $10 for the hour long study. Three reported limited prior experience of VR/AR. The procedure for each participant was identical: the study began with instructions and preliminary form filling followed immediately by setup of the equipment. They were then able to practice freely for a maximum of five minutes before beginning the formal study. In this way, we planned to retain 1944 correct tap trials (18 participants by three tap-techniques by two sizes by six positions by three repetitions) and, similarly, 1620 correct pan trials for analysis. A technical error led to the loss of one participant's data in the tap task, leaving 1836 tap trials for analysis.

Results and Discussion

Time and error data for tap and pan tasks are presented in Figures 4 and 5. All analyses were three-way repeated measure ANOVAs corrected for sphericity violations with Greenhouse-Geisser corrections and followed by *post-hoc* pairwise t-tests adjusted with Bonferroni confidence interval adjustments. For brevity, we report only significant ($\alpha < 0.01$) results - see Table 3 for interactions and main effects. *Post-hoc* comparisons are discussed in the following sections.

In the tap task, there was a single weak interaction effect. Accordingly, we opt to interpret the results in terms of the moderate to high power main effects. All three tap-techniques differed significantly in terms of the task time (all p<0.001). Unsurprisingly, the ability to make position adjustments with

Table 3. Significant RM ANOVA results from the EarTouch study.

Task	*Measure*	*Comparison*	*Outcome*		
Tap	Time	Tap-Technique	$F(2,32) = 80.6$	$p < 0.001$	$\eta_p^2 = 0.85$
		Size	$F(1,16) = 17.8$	$p = 0.001$	$\eta_p^2 = 0.53$
	Error Rate	Tap-Technique	$F(2,32) = 43.3$	$p < 0.001$	$\eta_p^2 = 0.73$
		Size	$F(1,16) = 173.5$	$p < 0.001$	$\eta_p^2 = 0.92$
		Tap-Tech. x Location.	$F(10,160) = 3.99$	$p < 0.001$	$\eta_p^2 = 0.2$
Pan	Time	Pan-Technique	$F(2,34) = 12.54$	$p < 0.001$	$\eta_p^2 = 0.42$
		Length	$F(4,68) = 121.94$	$p < 0.001$	$\eta_p^2 = 0.88$
	Error Rate	Length	$F(4,68) = 5.1$	$p = 0.001$	$\eta_p^2 = 0.23$

dwell and liftOff required time and, with liftOff, the need to trigger selection via an explicit finger-up event took yet longer. The benefits of these increased task times are clearly observed in the lower error rates for these conditions: both significantly improve over landOn error rates (p<0.001). In addition, and unsurprisingly, small targets also led to greatly increased error rates, and modestly increased task times, when compared to large targets. Data from the pan task were more uniform. The joystick pan-technique led to faster task times than both drag (p=0.002) and toggle (p=0.001) and task times, and to a lesser extent error rates, predictably increased with distance.

These results provide a window into comparing our system with prior work and deriving appropriate target sizes and techniques for ear based touch interfaces. In terms of the time data, figures for landOn data (1.48s) are relatively similar to those recorded on FaceTouch's [10] head mounted touch screen (1.39s, reported for touches to the side of the head). On the other hand, data from the LiftOff condition (2.5s) are noticeably slower than with FaceTouch (2.07s). This suggests that while initial touches can be readily performed, it may be more challenging to control a cursor via touches to the skin of the ear than via a standard touchscreen. A candidate explanation for this difference is that the skin is not as smooth, low-friction or uniform as a touchscreen - moving along its surface precisely is more difficult. In terms of errors, EarPut's [20] is the most directly comparable work. Its capacitive sensing wrap-around ear sensors logged 42% failures with six targets positioned on the edge of the ear, equivalent to the most error-prone landOn condition in the current study. In contrast, errors on large targets with the liftOff technique drop to a mean of 8.9%, a substantial improvement over this prior work. This suggests that optical tracking may offer considerable advantages over capacitive sensing for ear augmentation.

To better shed light on this issue, we examined the precision of both tap and pan inputs. For this analysis, we extracted the distance between the points selected and the target center in all trials (i.e., including errors) and processed it as follows. We discarded outliers more than three SD from the absolute mean, removing 93 (2.9%) trials from tap and 30 (1%) from pan. We then recalculated mean and SD values and report precision as mean plus/minus three times the SD (in normalized 0-100 units). This should account for 99.7% of the inputs intended to reach a given target location. Results for tap were: dwell

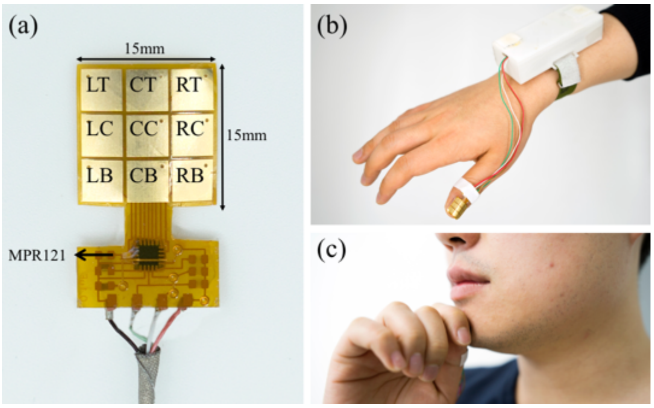

Figure 6. Prototype of the ThumbTouch system. (a) Dimension of the sensor array and labeling of target locations. (b) Wireless wearable design (c) Example use scenario.

(M:4.8, SD:4.8, precision:19.2); landOn (M:7.6, SD:6.0, precision:25.6) and; liftOff (M:4.3, SD:4.4, precision:17.5). This suggests liftOff and dwell will perform optimally with five targets, while landOn is more suited to a four target system. Data for pan are: drag (M:6.6, SD:9.9, precision:36.3); joystick (M:5.7, SD:8.8, precision:32.1) and; toggle (M:7.5, SD:11.6, precision:42.3). These values suggest that pan requires fewer and larger targets than tap for optimal input - three targets in the best performing joystick system. This suggests that pan input tasks for ear based systems may have limited expressivity. However, given the comparatively low error rates observed in the pan tasks (see Table 5), further study of pan input would be needed to confirm this recommendation.

Case Study 2 - ThumbTouch

The second case study explores input via a device mounted on the hand, rather than on the face. Specifically, we developed a capacitive sensing thumbnail, similar to Kao *et al.*'s NailO [15], and applied it to the previously unexamined setting of hand-to-face input – prior work has only considered finger touches to a sensing nail. This design was motivated by outcomes from the elicitation study. Specifically we envisaged nail touches to the cheek and chin, the two most frequently used face regions, and sought to embody three of the design strategies: *screening*, as touches to the face by the thumbnail could remain hidden behind the fingers; *re-purposing*, as such touches might support co-opting common actions such as scratching the cheek or gripping the chin and; *miniaturizing*, as input under such constraints is inevitably small in scale.

We developed a touch sensitive thumbnail by mounting a 0.3mm thick flexible PCB covered with a three-by-three grid of 4.5mm square electrodes, spaced with 0.4mm gaps, directly on the thumbnail. This device was intentionally designed to be much thinner (0.3mm) than prior systems that integrate all components into the nail (e.g., [15], 4mm). This is likely important in our scenario as the nail is used to touch rather than be touched by another finger. To minimize noise, the electrodes were connected to an MPR121 capacitive sensing micro-controller mounted just behind the nail on the thumb's distal phalanx. The MPR121 was connected to an Arduino

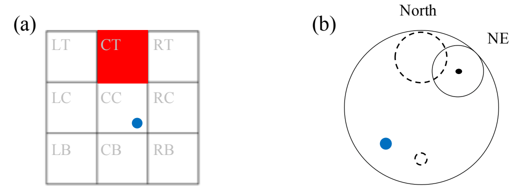

Figure 7. Interface for ThumbTouch study. (a) 3 by 3 grid used in tapping task showing target (red square) and cursor (blue dot). (b) Panning task interface for NE direction. Dotted line indicates another example: a North direction task.

Fio mounted on the wrist, which streamed data via a wireless xBee link to a host PC, which then processed this data and transferred it, via OSC, to the same BT-200 HMD used in the EarTouch study. All graphical interfaces were shown on this HMD. Sensor latency and update rate were approximately 7ms and 60Hz. The nail was attached to participants via double sided tape and the MPR121 chip secured with a band-aid. A key goal for this hardware design was to minimize the thickness of the nail sensor as preliminary testing during development indicated that a thicker nail would impede performance. The prototype is shown in Figure 6

We acquired touch position data from this sensor by following Oakley *et al.*'s [25] use of the ratio of baseline to currently measured capacitance on each MPR121 electrode to derive a three by three grid of touch intensities. We then processed this data using a three window median filter to remove noise, performed a bicubic upscale by a factor of three to increase resolution and extracted touch regions via Xiao *et al.*'s [42] method of flood fill to identify individual touches and image moments to calculate properties such as their centroid. As we were interested in a single touch point on the face, we always considered only the largest touch. As with EarTouch, we ignored data from the first and last 100ms of each touch.

Performance Study

The study broadly followed the format of the EarTouch study, adapted to the nail device format. We highlight key differences below. We again studied both tap and pan tasks. In the tap study, we maintained the three *tap-techniques* of *landOn*, *liftOff* and *dwell* and used nine 4.5mm square targets, arranged in a three-by-three grid that matched the sensor electrodes. These locations are identified by two-letter acronyms such as *LC* for Left-Center (see Figure 6.a). In addition, we studied thumbnail touches to two *face-sites*: the *chin* and *cheek*. This led to a 3 (tap-techniques) by 9 (target-locations) by 2 (face-sites) design. In the pan study, we used just two *pan-techniques*: *drag* and a combined *joystick*-toggle. We combined these techniques to better fit the small nail input device. The unified technique simply varied cursor speed proportionally (at twice displacement/second) to the distance of a touch from the nail's center point. Additional, we studied eight *pan-directions* (cardinal and semi-cardinal) and the chin and cheek *face-sites*. We used a single *pan-distance* of 10mm, always starting from one edge of the sensor/screen and moving across its center to the opposite side. Graphical feedback in these studies was updated to a 2D setting with targets and cursors shown as squares or circles (see Figure 7).

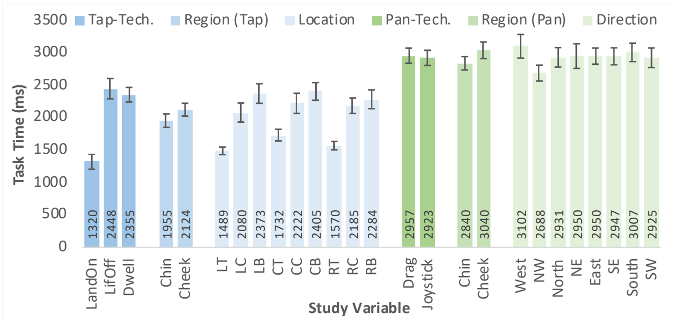

Figure 8. Task times in ThumbTouch tap/pan tasks. Bars show Std Err.

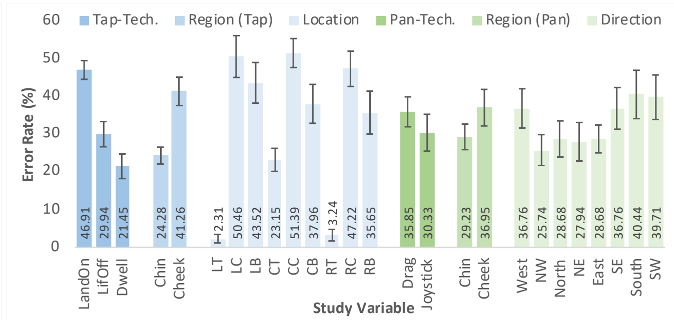

Figure 9. Error rates in ThumbTouch tasks. Bars show Std Err.

Table 4. Significant RM ANOVA results from the ThumbTouch study.

Task	*Measure*	*Comparison*	*Outcome*		
Tap	Time	Tap-Technique	F(2,34) = 26.8	$p < 0.001$	$\eta_p^2 = 0.63$
		Location	F(4.5,71.9) = 14.1	$p < 0.001$	$\eta_p^2 = 0.47$
		Tap-Tech. x Reg.	F(2,34) = 8.6	$p = 0.001$	$\eta_p^2 = 0.35$
		Tap-Tech. x Loc.	F(6.5,104.2) = 3.5	$p = 0.003$	$\eta_p^2 = 0.18$
	Error Rate	Tap-Technique	F(2,34) = 48.8	$p < 0.001$	$\eta_p^2 = 0.74$
		Region	F(1,17) = 28.1	$p < 0.001$	$\eta_p^2 = 0.62$
		Location.	F(4.6,77.8) = 27.7	$p < 0.001$	$\eta_p^2 = 0.62$
		Tap-Tech x Loc.	F(7.8,133.2) = 3.3	$p = 0.002$	$\eta_p^2 = 0.16$
		Region x Loc.	F(8,136) = 5.2	$p < 0.001$	$\eta_p^2 = 0.23$
Pan	Time	Pan-Tech. x Dir.	F(7,98) = 5.4	$p < 0.001$	$\eta_p^2 = 0.28$
	Error Rate	Pan-Tech. x Dir.	F(7,119) = 3.35	$p = 0.003$	$\eta_p^2 = 0.16$

Procedures followed the EarTouch study: participants were provided with instructions, donned the equipment and practiced freely for up to five minutes. They then completed tap followed by pan tasks in a repeated measures study design. In both tasks, tap/pan–technique and face–region were fully balanced, while target-location/–direction were randomized. A block was one set of directions or locations and participants completed three blocks per condition, with the first block considered practice and not retained for analysis. Pilot tests indicated that some input tasks were extremely challenging, so we opted not to require participants redo trials in which they made an error. In total 18 participants (10 males, mean age 24, all right-handed) completed this study. All were UNIST students, or recent graduates, and compensated with approximately $10 for the one hour experiment. 11 had limited prior experience of VR/AR. In total, we captured 1944 taps (18 participants by 3 tap-techniques by 2 face-sites by 9 target-locations by 2 blocks) and 1152 pans (18 participants by 2 pan-techniques by 2 face-sites by 8 pan-directions by 2 blocks).

Results and Discussion

We analyzed results using similar methods to the EarTouch study: repeated measures ANOVA following by *post-hoc* testing, with sphericity and confidence interval adjustments applied where appropriate. We report only significant results ($\alpha < 0.01$). Time and error data are depicted in Figures 8 and 9 and ANOVA results are shown in Table 4. Time data includes measurements from both successful and error trials as we observed few differences between the aggregate performance of these sets (overall means and SDs were 38-114ms apart) and, given failed trials were not repeated, the entire set was considerably more complete. Data from timeouts are not shown (38 or 1.95% of tap trials and 31 or 2.7% of pan).

Performance in the tap task varied considerably in both time and accuracy. As in the EarTouch study, landOn was both faster and more error prone than both liftOff and Dwell (all p<0.001): without interactive feedback, the tap task can be executed quickly, but not accurately. Touches to the cheek also led to significantly more errors, but not longer times, than touches to the chin. We speculate this is due to differences in compliance between the relatively rigid chin and soft cheek - it was more difficult to make accurate taps on a soft skin surface. Due to the large number of comparisons it involves, we opted not to conduct pairwise tests on the location variable; instead we depict this data in the confusion matrix in Figure 10. The error interaction effects (not plotted) are due to the accurate performance of the TL and TR locations (the left and right tips of the nail) remaining unaffected by the main effects of technique and region: these were readily accessible locations. The interaction of technique by region in the time data suggests dwell is faster on the chin that the cheek, while the relatively weak technique by location effect had no clear interpretation. Performance in the more challenging pan task was more uniform and the two weak interaction effects suggest that the joystick technique performed better than the drag technique for upward directions (NW, north and NE). This again reflects the increased accessibility of these regions - it is easier to touch the top of the nail.

The tap confusion matrix sheds more light on this issue. While the strong performance of the TL and TR tip locations is inarguable, we also note that nail edges may be reliably detectable using a more advanced algorithm - touches to the LC, TC and RC are correctly associated with the left, top and right of the nail in a mean of 96.9% of cases. This indicates participants were able to select the correct side of the nail, but had difficulty selecting specific locations on that side. This is likely due the whole edge of the nail making contact with the face. We speculate that side touches could be accurately performed and easily disambiguated from tip touches by examining the size of the contact area. Similarly, we note that the CC and CB touches are the only two locations with errors distributed over all other locations. This suggests that the input action was akin to simply pressing the whole nail against the face, an action that could also likely be distinguished from edge contacts by examining touch region sizes [25]. Adding this functionality and capturing data on these inputs is a clear next step for this work.

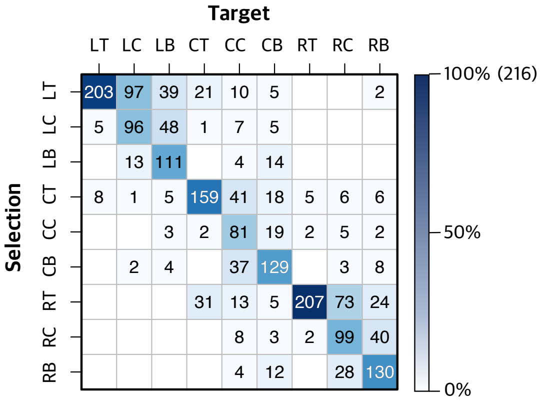

Figure 10. Confusion matrix for ThumbTouch tap task.

We also calculated precision data. After discarding outliers (tap: 64 or 2.1%, pan: 72 or 4%), the values (in mm) were: dwell (M:1.94, SD:1.26, precision:5.72); landOn (M:3.15, SD:1.62, precision:8.01); liftOff (M:2.45, SD:1.53, precision:7.04); drag (M:2.16, SD:1.62, precision:7.02) and; joystick (M:2.88, SD:2.95, precision:11.73). These relatively large figures, relative to sensor size, confirm the difficulty of tasks studied and the need to consider alternative approaches.

VALIDATION STUDY

While the lab studies of the Ear– and ThumbTouch extend prior work and contribute practical assessments of performance in hand-to-face input tasks, they do not address our core goal of improving our understanding of socially acceptable hand-to-face input. Accordingly, we conducted a final study to assess the input techniques, and the design guidance they instantiate, from the perspectives of social acceptability and subtle or non-obvious/unobtrusive input. In this study participants operated or observed the HMD and Ear/Thumb prototypes in representative input tasks in the same coffee shop used in the original elicitation study. In total, 12 new participants (4 males, mean age 23.7, all right-handed, all UNIST students) were compensated with $10 to complete the study. As in the elicitation study, they worked in six pairs (all strangers).

One participant in each pair took the role of the *user*, wearing and operating the devices, while the other acted as an *observer*, watching these activities, but not briefed in advance that input actions were taking place. The study presented three input tasks on both devices. On EarTouch, participants experienced the liftOff (tap), drag and joystick (both pan) tasks, while on ThumbTouch they experienced dwell (tap), drag and joystick (both pan). These techniques were selected to ensure a diverse set of techniques and/or due to their comparatively strong performance in the lab studies. The order the prototypes were experienced was balanced between participant pairs, while the order of three techniques on each prototype was varied using a Latin square design. For each task/device pair, the user participants performed the training tasks from the original studies, experiencing the input and feedback (but not performing any sustained, repetitive input tasks) while the observer watched. Between each task, both participants filled out the social acceptability/obviousness questionnaires from the elicitation study. At the end of the study, we conducted a semi-structured interview with the pair of participants capturing opinions and reactions to the input techniques. Participants were asked to reflect on and contrast the techniques in terms of general opinions and how comfortable and unobtrusive (or uncomfortable and obvious) they felt they were. Finally, we note that due to problems with the lighting conditions in the coffee shop, the EarTouch system performed poorly for three participants. These participants received an explanation of the system and were able to partially experience it, but the reliability of the input was low. This may have influenced performed actions, opinions and ratings.

Experiment Results

The data from the questionnaires are summarized, in the form of the compound acceptability scores we used in the elicitation study, in Table 5. The data show that *users* tended to rate tap tasks (liftOff/Dwell) as socially acceptable and unobtrusive in more locations and situations than pan tasks (drag/joystick). No such clear cut effect was observed in the the data from the *observers*. Furthermore, they tended to score lower than the *users*. The likely reflects the fact that the *observers* were naïve. The input actions, without being situated in the context of controlling an interactive system, may have appeared less acceptable than to the fully-aware *users* operating them. To shed further light on these issues, we turned to the diverse set of comments and opinions captured in the interviews, which we processed by transcribing them and then organizing them via an iterative affinity process. In this description, quotes are marked with *U* for device *Users* and an *O* for *Observers*.

We first focused on comments with respect to the five design strategies from the elicitation study. All participants commented on the value of the *camouflaging* strategy. Eight participants viewed the ear as a suitable site for masking input actions as it was commonly touched. It was "OK because it is similar to touching an earring" (U1), or things "people already wear in their ear like Bluetooth earphones" (O4). Two participants (U6, O1) also referenced sweeping back their hair over their ears. O4 suggested that "even if there was a stranger passing by they would never know" the touches were controlling a device and O2 remarked that, if s/he had not been participating in a study, s/he "would not have thought of assessing or paying attention to [the ear touches] at all". Five participants felt touches to the chin were also "natural" (U1) or "habitual" (O4) and touching it resembles common behavior such as "thinking or when they scratch it" (O2, O6) or that it was simply a region where people usually touch their faces (O5). Participants also offered contrasting opinions: the ear "stands out too much" compared to the chin (O2) or that people "do not normally touch their chins" (O3) – it is an "[un]common form" of touch (U3). Ultimately, the diversity in this assessment reflects what

Table 5. Mean acceptability scores (0-1) in the validation study.

Role	*Prototype*	*Overall*	*LiftOff/Dwell*	*Drag*	*Joystick*
User	EarTouch	0.74	0.85	0.67	0.70
	NailTouch	0.70	0.81	0.70	0.60
Observer	EarTouch	0.58	0.57	0.65	0.51
	NailTouch	0.69	0.68	0.75	0.64

O3 referred to as the "gestures that we normally do as habit" – to be successful, camouflage needs to adapt an existing user behavior, a "daily gesture like touching or putting" (U1). It thus varies from user to user.

The *miniaturizing* strategy was highlighted as valuable by nine participants. Two felt the control actions were not "big, [so] they were mostly OK" (O1) or were simply "too small to notice" (U4), while one emphasized the importance of touches to a "small region" (O3). Others appreciated ThumbTouch as its "small movements won't be weird" (U2) or favored it over EarTouch as "the gesture for the ear was too big up and down" (O5) and you have to raise "your arm higher to reach the ear" (O6). The value of *miniaturizing* also came across in an assessment of the input techniques – the larger or prolonged movements in the pan tasks were "weird" and would "gather the attention" (U1) and actions that involved larger physical movements such as "taking the finger off the ear" in liftOff (O2) stood out. Although EarTouch involved larger movements, it benefited from the *obfuscating* strategy of using the head to obscure input. O1 remarked on the "difference between front and side" and touches to the ear being less "intrusive" that those on the chin, while O4 acknowledged that "the active range was much bigger for ear gestures, but since the ear is on the side, it didn't matter". Participants also suggested this strategy for new designs: under the chin, on the neck or behind the ear would "hide the gesture well" (O6). The *screening* strategy was mentioned by a single participant who noted that in ThumbTouch they "could hide my thumb so I can make others not notice this" (U1), while the *re-purposing* strategy did not emerge in the interviews, possibly due to its relatively weak embodiment in the two prototypes.

Beyond these comments on the strategies, participants brought up a range of other issues. U6 was concerned about hand-to-face touches and makeup, an issue that has cropped up in prior work [31], while U5 appreciated that neither input technique required a hand-held controller and also the proprioceptive aspects of the tasks: it was convenient that s/he "could know where the hand is touching now". In general, participants, and particularly four from the six observers, were positive about performing the input actions in the coffee shop setting. O3 remarked that s/he "didn't think people would care about the gestures in a cafe", O4 that they "certainly didn't bother me" and O6 that the input actions "are fine to do" as a cafe as it is "a place where people move around and talk". In addition, O6 felt they would be suitable "on the street or on the bus", while O5 was more cautious: in a cafe, "it was ok", but somewhere "quiet and static it would have stuck out".

The results from this study generally validate the design strategies identified in the elicitation study and instantiated in the prototypes. In particular, the interview data reveal that both users and observers, who respectively conducted and naïvely viewed input tasks on the prototypes, frequently highlighted aspects of the strategies as important qualities or mechanisms to achieve socially acceptable and non-obvious or unobtrusive hand-to-face input. For both prototypes, strategies such as camouflaging and miniaturizing were viewed as critical, whereas both obfuscating and screening were viewed as effective aspects of the techniques by smaller numbers of participants. A key topic for future work on this topic would be to tackle issues of how unobtrusiveness could be sustained for prolonged or frequent use. One way to achieve this might be to provide multiple input mechanisms [2], so that users would be able to vary how they performed input depending on the situation they were in, rather than engage in sequences of repetitive motions.

CONCLUSIONS

This paper describes a multi-stage research process that identifies and validates design strategies for making subtle, unobtrusive and socially acceptable hand-to-face input. Specifically, from an elicitation study focused on social acceptability we derive, and contribute, five strategies for the design of socially acceptable hand-to-face input techniques. We instantiate these strategies in two novel prototype input systems and contribute empirical characterizations of their use, extending knowledge about human performance in hand-to-face interaction. Finally, we validate the design strategies in a study that has user and observer participants experience the prototypes in a public setting. We reflect on the importance of the different strategies in achieving socially acceptable hand-to-face input.

Limitations of this work include a reliance on participants from a single culture and age range; more diverse participants would improve its validity. The current study could also have been subject to experimenter effects [40] and follow up work should seek to isolate their impact by, for example, objectively logging bystander reactions in response to hand-to-face input. Additionally, we note the current prototypes do not represent a comprehensive exploration of the design space outlined by the design strategies; additional cases should be developed to better showcase the value of the strategies. The immaturity of the prototypes (e.g., exposed wires, visible cameras) may also have impacted the results, although we expect this would only have lowered perceptions of social acceptability and that more mature devices would only boost these ratings. Indeed, the prototypes could be improved in many ways. Next steps for EarTouch involve integrating machine learning algorithms to detect a greater number of hand actions, such as bends, pushes or deformations of the ear. For ThumbTouch, a clear follow-up is to create a system focused on capturing performance using edges of the nail rather than 2D positions and on enhancing performance in directional input tasks. One possibility here would be to focus on open-loop tasks such as swipe. Finally, longer term field studies of hand-to-face input will be required to move beyond some of the reservations advanced by our participants: the techniques described in this paper were deemed appropriate for short term use in public settings such as coffee shops, but over more prolonged periods, the input techniques they feature, and strategies they represent, may not hold up to scrutiny. Future work should look into deploying hand-to-face input systems to participants for sustained use to better understand these long-term, real-world, effects.

ACKNOWLEDGMENTS

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (2017R1D1A1B03031364).

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 (MobileHCI '14)*. ACM, New York, NY, USA, 193–202. DOI: http://dx.doi.org/10.1145/2628363.2628381
2. Fraser Anderson, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2015. Supporting Subtlety with Deceptive Devices and Illusory Interactions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1489–1498. DOI: http://dx.doi.org/10.1145/2702123.2702336
3. Ronald T. Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoper. Virtual Environ.* 6, 4 (Aug. 1997), 355–385. DOI: http://dx.doi.org/10.1162/pres.1997.6.4.355
4. Liwei Chan, Yi-Ling Chen, Chi-Hao Hsieh, Rong-Hao Liang, and Bing-Yu Chen. 2015. CyclopsRing: Enabling Whole-Hand and Context-Aware Interactions Through a Fisheye Ring. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 549–556. DOI: http://dx.doi.org/10.1145/2807442.2807450
5. Enrico Costanza, Samuel A. Inverso, and Rebecca Allen. 2005. Toward Subtle Intimate Interfaces for Mobile Devices Using an EMG Controller. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 481–489. DOI:http://dx.doi.org/10.1145/1054972.1055039
6. Christine Dierk, Sarah Sterman, Molly Jane Pearce Nicholas, and Eric Paulos. 2018. HäirIÖ: Human Hair As Interactive Material. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '18)*. ACM, New York, NY, USA, 148–157. DOI: http://dx.doi.org/10.1145/3173225.3173232
7. Stuart Dimond and Rashida Harries. 1984. Face touching in monkeys, apes and man: Evolutionary origins and cerebral asymmetry. *Neuropsychologia* 22, 2 (1984), 227 – 233. DOI:http://dx.doi.org/https://doi.org/10.1016/0028-3932(84)90065-4
8. David Dobbelstein, Philipp Hock, and Enrico Rukzio. 2015. Belt: An Unobtrusive Touch Input Device for Head-worn Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2135–2138. DOI: http://dx.doi.org/10.1145/2702123.2702450
9. Barrett Ens, Ahmad Byagowi, Teng Han, Juan David Hincapié-Ramos, and Pourang Irani. 2016. Combining Ring Input with Hand Tracking for Precise, Natural Interaction with Spatial Analytic Interfaces. In *Proceedings of the 2016 Symposium on Spatial User Interaction (SUI '16)*. ACM, New York, NY, USA, 99–102. DOI: http://dx.doi.org/10.1145/2983310.2985757
10. Jan Gugenheimer, David Dobbelstein, Christian Winkler, Gabriel Haas, and Enrico Rukzio. 2016. FaceTouch: Enabling Touch Interaction in Display Fixed UIs for Mobile Virtual Reality. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 49–60. DOI: http://dx.doi.org/10.1145/2984511.2984576
11. Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: Appropriating the Body As an Input Surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 453–462. DOI: http://dx.doi.org/10.1145/1753326.1753394
12. Takeshi Hatta and Stuart J. Dimond. 1984. Differences in face touching by Japanese and British people. *Neuropsychologia* 22, 4 (1984), 531 – 534. DOI: http://dx.doi.org/https://doi.org/10.1016/0028-3932(84)90050-2
13. Yi-Ta Hsieh, Antti Jylhä, Valeria Orso, Luciano Gamberini, and Giulio Jacucci. 2016. Designing a Willing-to-Use-in-Public Hand Gestural Interaction Technique for Smart Glasses. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4203–4215. DOI: http://dx.doi.org/10.1145/2858036.2858436
14. Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E. Froehlich. 2015. Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 241–250. DOI: http://dx.doi.org/10.1145/2702123.2702393
15. Hsin-Liu (Cindy) Kao, Artem Dementyev, Joseph A. Paradiso, and Chris Schmandt. 2015. NailO: Fingernails As an Input Surface. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3015–3018. DOI: http://dx.doi.org/10.1145/2702123.2702572
16. Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1151–1160. DOI: http://dx.doi.org/10.1145/2638728.2641695

17. 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 (MobileHCI '17)*. ACM, New York, NY, USA, Article 27, 6 pages. DOI: http://dx.doi.org/10.1145/3098279.3098538

18. Yen Lee Angela Kwok, Jan Gralton, and Mary-Louise McLaws. 2015. Face touching: A frequent habit that has implications for hand hygiene. *American Journal of Infection Control* 43, 2 (2015), 112 – 114. DOI: http://dx.doi.org/https://doi.org/10.1016/j.ajic.2014.10.015

19. Juyoung Lee, Hui-Shyong Yeo, Murtaza Dhuliawala, Jedidiah Akano, Junichi Shimizu, Thad Starner, Aaron Quigley, Woontack Woo, and Kai Kunze. 2017. Itchy Nose: Discreet Gesture Interaction Using EOG Sensors in Smart Eyewear. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers (ISWC '17)*. ACM, New York, NY, USA, 94–97. DOI: http://dx.doi.org/10.1145/3123021.3123060

20. 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 (OzCHI '14)*. ACM, New York, NY, USA, 300–307. DOI: http://dx.doi.org/10.1145/2686612.2686655

21. Calkin S. Montero, Jason Alexander, Mark T. Marshall, and Sriram Subramanian. 2010. Would You Do That?: Understanding Social Acceptance of Gestural Interfaces. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '10)*. ACM, New York, NY, USA, 275–278. DOI: http://dx.doi.org/10.1145/1851600.1851647

22. Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel, and Jacob O. Wobbrock. 2014. Reducing Legacy Bias in Gesture Elicitation Studies. *interactions* 21, 3 (May 2014), 40–45. DOI:http://dx.doi.org/10.1145/2591689

23. Christian Müller-Tomfelde. 2007. Dwell-based Pointing in Applications of Human Computer Interaction. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-computer Interaction (INTERACT'07)*. Springer-Verlag, Berlin, Heidelberg, 560–573. http://dl.acm.org/citation.cfm?id=1776994.1777067

24. Mark Nicas and Daniel Best. 2008. A Study Quantifying the Hand-to-Face Contact Rate and Its Potential Application to Predicting Respiratory Tract Infection. *Journal of Occupational and Environmental Hygiene* 5, 6 (2008), 347–352. DOI: http://dx.doi.org/10.1080/15459620802003896 PMID: 18357546.

25. Ian Oakley, Carina Lindahl, Khanh Le, DoYoung Lee, and MD. Rasel Islam. 2016. The Flat Finger: Exploring Area Touches on Smartwatches. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4238–4249. DOI: http://dx.doi.org/10.1145/2858036.2858179

26. Masa Ogata, Yuta Sugiura, Yasutoshi Makino, Masahiko Inami, and Michita Imai. 2013. SenSkin: Adapting Skin As a Soft Interface. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 539–544. DOI: http://dx.doi.org/10.1145/2501988.2502039

27. Iulian Radu. 2014. Augmented Reality in Education: A Meta-review and Cross-media Analysis. *Personal Ubiquitous Comput.* 18, 6 (Aug. 2014), 1533–1543. DOI: http://dx.doi.org/10.1007/s00779-013-0747-y

28. Jun Rekimoto. 2001. GestureWrist and GesturePad: Unobtrusive Wearable Interaction Devices. In *Proceedings of the 5th IEEE International Symposium on Wearable Computers (ISWC '01)*. IEEE Computer Society, Washington, DC, USA, 21–. http://dl.acm.org/citation.cfm?id=580581.856565

29. Julie Rico and Stephen Brewster. 2010. Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 887–896. DOI: http://dx.doi.org/10.1145/1753326.1753458

30. Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined Motion Gestures for Mobile Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 197–206. DOI: http://dx.doi.org/10.1145/1978942.1978971

31. 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 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3181–3190. DOI: http://dx.doi.org/10.1145/2556288.2556984

32. Outi Tuisku, Ville Rantanen, Oleg Åăpakov, Veikko Surakka, and Jukka Lekkala. 2016. Pointing and Selecting with Facial Activity. *Interacting with Computers* 28, 1 (2016), 1–12. DOI: http://dx.doi.org/10.1093/iwc/iwu026

33. Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1325–1334. DOI: http://dx.doi.org/10.1145/2702123.2702223

34. G. Stewart Von Itzstein, Mark Billinghurst, Ross T. Smith, and Bruce H. Thomas. 2017. *Augmented Reality Entertainment: Taking Gaming Out of the Box*. Springer International Publishing, Cham, 1–9. DOI: http://dx.doi.org/10.1007/978-3-319-08234-9_81-1

35. Cheng-Yao Wang, Wei-Chen Chu, Po-Tsung Chiu, Min-Chieh Hsiu, Yih-Harn Chiang, and Mike Y. Chen. 2015. PalmType: Using Palms As Keyboards for Smart Glasses. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, New York, NY, USA, 153–160. DOI: http://dx.doi.org/10.1145/2785830.2785886

36. Feng Wang and Xiangshi Ren. 2009. Empirical Evaluation for Finger Input Properties in Multi-touch Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1063–1072. DOI: http://dx.doi.org/10.1145/1518701.1518864

37. J. Nancy Wei, Bryn Dougherty, Aundria Myers, and M. Sherif Badawy. 2018. Using Google Glass in Surgical Settings: Systematic Review. *JMIR Mhealth Uhealth* 6, 3 (06 Mar 2018), e54. DOI: http://dx.doi.org/10.2196/mhealth.9409

38. Martin Weigel, Vikram Mehta, and Jürgen Steimle. 2014. More Than Touch: Understanding How People Use Skin As an Input Surface for Mobile Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 179–188. DOI: http://dx.doi.org/10.1145/2556288.2557239

39. Martin Weigel, Aditya Shekhar Nittala, Alex Olwal, and Jürgen Steimle. 2017. SkinMarks: Enabling Interactions on Body Landmarks Using Conformal Skin Electronics. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3095–3105. DOI: http://dx.doi.org/10.1145/3025453.3025704

40. Julie R. Williamson and John Williamson. 2017. Understanding Public Evaluation: Quantifying Experimenter Intervention. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3414–3425. DOI: http://dx.doi.org/10.1145/3025453.3025598

41. Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1083–1092. DOI: http://dx.doi.org/10.1145/1518701.1518866

42. Robert Xiao, Julia Schwarz, and Chris Harrison. 2015. Estimating 3D Finger Angle on Commodity Touchscreens. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces (ITS '15)*. ACM, New York, NY, USA, 47–50. DOI: http://dx.doi.org/10.1145/2817721.2817737

43. Koki Yamashita, Takashi Kikuchi, Katsutoshi Masai, Maki Sugimoto, Bruce H. Thomas, and Yuta Sugiura. 2017. CheekInput: Turning Your Cheek into an Input Surface by Embedded Optical Sensors on a Head-mounted Display. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17)*. ACM, New York, NY, USA, Article 19, 8 pages. DOI: http://dx.doi.org/10.1145/3139131.3139146