Exploring User Defined Gestures for Ear-Based Interactions

YU-CHUN CHEN, National Chiao Tung University, Taiwan and National Taiwan University, Taiwan
CHIA-YING LIAO, National Chiao Tung University, Taiwan
SHUO-WEN HSU, National Chiao Tung University, Taiwan
DA-YUAN HUANG, Human Machine Interaction Lab, Huawei, Canada
BING-YU CHEN, National Taiwan University, Taiwan

The human ear is highly sensitive and accessible, making it especially suitable for being used as an interface for interacting with smart earpieces or augmented glasses. However, previous works on ear-based input mainly address gesture sensing technology and researcher-designed gestures. This paper aims to bring more understandings of gesture design. Thus, for a user elicitation study, we recruited 28 participants, each of whom designed gestures for 31 smart device-related tasks. This resulted in a total of 868 gestures generated. Upon the basis of these gestures, we compiled a taxonomy and concluded the considerations underlying the participants' designs that also offer insights into their design rationales and preferences. Thereafter, based on these study results, we propose a set of user-defined gestures and share interesting findings. We hope this work can shed some light on not only sensing technologies of ear-based input, but also the interface design of future wearable interfaces.

CCS Concepts: • **Human-centered computing → Gestural input**; User studies.

Additional Key Words and Phrases: Gestures; Ear-based Input; User-defined; Guessability; Think-aloud.

ACM Reference Format:
Yu-Chun Chen, Chia-Ying Liao, Shuo-wen Hsu, Da-Yuan Huang, and Bing-Yu Chen. 2020. Exploring User Defined Gestures for Ear-Based Interactions. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 4, ISS, Article 186 (November 2020). ACM, New York, NY, USA. 20 pages. https://doi.org/10.1145/3427314

1 INTRODUCTION

Thanks to the advances of technologies, wearable devices have become compact yet powerful. However, in consideration of the wearer's physical and social comfort, such devices remain small limiting the ability of users to input interactions. To address this issue, researchers have proposed instrumentalizing areas on the human body (*e.g.*, the finger, limbs, or the face) for use as interactive spaces, as the skin and its surrounding space provide a larger input area and are very easily accessible. In addition, such interfaces allow users to communicate with wearable devices without the necessity of sight, in an "eyes-free" manner [14, 16, 18], making them especially usable in mobile scenarios.

Authors' addresses: Yu-Chun Chen, lesleychen@ntu.edu.tw, National Chiao Tung University, Hsinchu, Taiwan, National Taiwan University, Taipei, Taiwan; Chia-Ying Liao, npc90039.aa05g@nctu.edu.tw, National Chiao Tung University, Hsinchu, Taiwan; Shuo-wen Hsu, shuowen0619.aa06g@nctu.edu.tw, National Chiao Tung University, Hsinchu, Taiwan; Da-Yuan Huang, dayuan.huang@huawei.com, Human Machine Interaction Lab, Huawei, Ontario, Canada; Bing-Yu Chen, robin@ntu.edu.tw, National Taiwan University, Taipei, Taiwan.

2573-0142/2020/11-ART186 $15.00
https://doi.org/10.1145/3427314

Among many areas of the body, the human ear is appealing for use as seen from commercially popular, head-worn devices, such as smart earpieces or augmented glasses. Compact and integrated devices are possible because the ears are close to these devices and the area on or around the ears are used as input space. Furthermore, the human ear contains rich form factors. Previous works have shown that rich, eyes-free mobile interactions can be enabled by the use of touching the ears, applying midair gestures around the ears [18], or deforming the auricles [14]. They also showed that human proprioception enables users to have on-body interactions without a visual interface. The ears and fingers provide natural tactile feedback for touch-based input, allowing for eyes-free interactions that do not interfere with social activities. The ears are easily accessible for single-handed or bimanual interaction. These studies made good use of the ear's specific affordances and focused on hardware development. Nevertheless, there is limited research that focuses on exploring intuitive gestures for ear-based input. To better understand how users' physical- and social-comfort is impacted when interacting with smart devices via ear-based input, we conducted a user elicitation study from which results we compile a set of user-defined gestures.

We correlated 31 frequently used tasks on smart devices. These tasks include visual-necessary tasks for which users need to see the target when executing the tasks, and non-visual-necessary tasks that users can perform without visual feedback. We recruited 28 participants to design unique gestures for every task without any constraint on the interaction method, such as gestures with or without touching or even deforming (*i.e.*, physically manipulating the shape of the ear by touch) the ear. They also rated their gestures on a continuous numeric rating from 1 to 7, with 1 indicating "strongly disagree" and 7 "strongly agree", according to suitability, usability, and social comfort. In summary, $31 \times 28 = 868$ gestures were gathered together in the elicitation study. We collected data and analyzed all gestures and subjective feedback through video records and transcripts of post-study interviews. We then compiled a taxonomy of the gestures and propose a set of user-defined gestures for ear-based input space. The design rationales, preferences, and the insights gained from this study are also discussed.

This study's contributions include: (1) a structured exploration of ear-based input by conducting a user elicitation study, (2) a taxonomy analysis to better understand how participants map versatile input methods for certain types of tasks, (3) a user-defined gesture set that allows future researchers to consider the sensing technologies needed for their devices, and (4) the observations and discussions from participants' feedback during the study.

2 RELATED WORK

2.1 On-Body Input

With our environment full of computer technology, we can often find ourselves in scenarios in which we cannot, or prefer not, to interact with a physical device in hand. Thus, many researchers have contributed to improve technology and discover new implementations to support device-free interaction. Saponas *et al.* [30] uses a forearm EMG to detect human muscular movement in real-time, and make finger interactions possible. Later, Matthies *et al.* [19] also proposed Botential, which can recognize the location of the tap gesture on any point of the entire body through EMG. Moreover, SkinTrack [49] is a wearable system implemented through emission of a continuous high-frequency AC signal, and a sensing wristband with multiple electrodes that enables continuous touch tracking on the skin. OmniTouch [8] is a wearable depth-sensing and projection system that enables users to expand the interactive areas on any surface of the users' body. The Magic Finger [47] uses an optical mouse sensor and a micro RGB camera to similarly afford touch input anytime and anywhere through a small device worn on a finger.

In consideration of social acceptance, Rekimoto *et al.* [27] created GestureWrist and GesturePad that were designed to be unobtrusive and used within various social contexts. GestureWrist is a wristband-type input device that can recognize forearm movements and hand gestures, and all sensing elements are hidden in a normal wristband. GesturePad is a sensing module which can be attached to the inside of clothes and allows users to interact with it from the outside. When it comes to microgesture, DigitSpace [9] utilizes thumb-to-fingers interfaces, which support one-handed and eyes-free input while wearing touch widgets on the fingers, and enables easy access via a stylus thumb. NailO [11] is an input device in the form of a nail sticker that works as a trackpad that can be customized and is removable exhibiting a good combination of on-body and fashion accessories.

Sensing technology at and around the ear has been explored in previous research, *e.g.*, EarField-Sensing [20] and CanalSense [1] present a facial-gesture and head-gesture recognition system with an in-ear placed sensing system. Ashbrook *et al.* [3] present instrumentation behind the ears that uses a bone-conductive microphone to sense bites. Several approaches have been proposed to enable on-body input at different points on the body. Although these alternatives were designed with the users' needs in mind, they mainly focus on the sensing technologies of the system.

2.2 Interaction with the Ears

Earphones are widely used in our daily lives due to their small size and wearability. Some commercial earphones are not only being used in a role as a music player but also have adopted the use of sensors that allow users to operate connected smartphones in a hands-free manner, such as the Sony Xperia Ear and the Apple Airpods. Some previous works have focused on interactions around the ear. Schwarz *et al.* [32] manipulates an MP3 player by twists and tugs on the headphone cord, making it into an input device. FreeDigiter [22] is an infrared proximity sensor with a dual-axis accelerometer which is small enough to be an in-ear hearing aid, and enables both recognition of hover finger gestures and the control of mobile devices. Lissermann *et al.* [18] presents EarPut, a novel interface concept and hardware prototype which instrumentalizes the human ear as an interactive surface by placing unobtrusive accessories behind the ear. Concerning the flexibility of the ear, Kikuchi *et al.* [14] proposes EarTouch, a new input method using several optical sensors attached to earphones to detect deformations of the ear. Lee *et al.* [16] further explores the social acceptability of facial touches and found that the ear garners a high score in terms of the facial region and social acceptability. These previous studies show how interaction with the ears is possible from mainly a technical perspective. In our study, we have attempted to discover what kind of input gestures are suitable for users and perceive their mental models.

2.3 User Elicitation for Gesture Design

Gesture-based interfaces are common in a variety of application domains, such as for desktops, mobile devices, augmented, virtual reality and gaming [12]. Initially for ear-based input, the gestures have been usually defined by researchers and limited by the sensing limitations of prototype devices. Some prior works directly employed users to define input systems, allowing users to join in the design process. This process is known as "participatory design" [31]. Nielsen *et al.* [25] describe a similar approach. They proposed a procedure that is useful for finding user-defined gestures that emphasizes the importance of using theory from ergonomics. User elicitation has also been used to design gesture interface on surface computing [23, 44], finger and hand motion gestures for control of TVs [36, 37, 40], hand gestures for augmented reality [26], gestures while using skin as an input surface for mobile computing [42], and when exploring different interactions between users and smart devices, such as mobile phones [29, 50], smart glasses [35], smartwatches [2] and other wrist-worn devices [13]. Villarreal-Narvaez *et al.* [41] provided a wide discuss about 216 exited user gesture elicitation studies.

When it comes to user-specified and expert-specified gestures, Wobbrock *et al.* find that gestures created by users were easier to master [23]. Furthermore, Nacenta *et al.* find that user-defined gestures are more memorable [24]. Rico and Brewster investigated possible ways of measuring social acceptability [28] and show that location and audience have a significant impact on a user's willingness to perform gestures. Also, to gain some insight into the user's mental model and to guide the designer in understanding the user's design space, numerous studies have classified gesture preferences into taxonomies based on gestures' characteristics, *e.g.*, Wobbrock *et al.* proposed a taxonomy which involves gestures for interactive tabletops [44]. Since then, many studies have adopted or extended the taxonomies for interaction to mobile phones [29], wearables [2, 33, 35], on-skin microgestures [5], augmented reality [26], and so on. Moreover, to analyze and interpret elicited data, Wobbrock *et al.* created a formula for agreement scores [43], which has been widely used by prior elicitation studies [26, 29, 35]. Later, Vatavu and Wobbrock modified this formula [39] to make it more accurately to represent findings. And Vatavu *et al.* [38] further provided a new approach to computing objective measures of consensus for the user gesture elicitation study. In this study, we used the user-elicitation methodology to identify users' expectations and suggestions for smart earpieces and then use the agreement rate proposed by Vatavu and Wobbrock [39] to evaluate our user-defined gesture set.

3 EXPLORING USER-DEFINED GESTURE FOR EARPIECE

In order to fully explore user-defined gestures for ear-based input, we elicited input from 28 participants. Participants were asked to design and perform a gesture around the ear that could execute or trigger a specific task for ear-based input. To reduce fatigue, participants designed only one unique gesture for each task. Note that no repetitive gestures were allowed for different tasks since one gesture cannot result in different outcomes without causing a conflict for the system.

As Lissermann *et al.*'s study [18] indicates that because of the small area available for touch on the ear, users like to use a variety of other atomic interaction primitives for ear-based input, in addition to touch-based gestures. Also, owing to the advancement of ear-based input recognition technology, many systems now can detect different interaction methods such as hovering hand gestures [22] and touch-based gestures [14, 18]. Therefore, to explore which interaction methods and gestures users intuitively performed for different tasks, we did not constrain the users' designing or reference any particular sensing technology. Instead, we sought to remove the gulf of execution [10] between the user's psychological goal and physical action. Thus, we encouraged participants to focus on gesture design and assume that all gestures are recognizable, *i.e.*, able to be perceived by some sensing technology. There is no recognition feedback, nor restrictions, provided during the performance of gestures, except that participants could only use one hand to perform the gesture and the interaction space was required to be near or on the ear.

Each participant received 31 tasks to execute during the study (see Table 1) and these took approximately 1.5 hours to complete. The individual gestures designed by participants for each task were extracted and labeled from the recorded video and transcripts. Based on the collected data, we conducted a taxonomical analysis and calculated the *agreement rate* (*AR*) of the gestures [39]. Ultimately we have proposed a user-defined gesture set.

3.1 Tasks

To understand how users interact with smart devices through ear-based input, we generated 31 tasks that are common on smartphones and computers, as head-worn accessories are commonly used to interact with them, as shown in Table 1. These tasks were selected from prior elicitation studies that focus on gestures for interactive surfaces [6, 29, 36, 40, 44] and wearable devices [2]. They include visual-necessary tasks for which users need to observe the target when executing

Table 1. The task list which is based on previous studies and grouped by category is presented to participants. There are two categories: navigation and action, and two subcategories of each: system and application.

Navigation		**Action**	
System	Application	System	Application
Go to Home Screen [2, 29]	Pan Right [2, 6, 29, 44]	Voice Search [29]	Turn on Microphone [2]
App Switch Next [2, 29]	Pan Left [2, 6, 29, 44]	Act on Selection [2, 6, 29, 44]	Turn off Microphone [2]
App Switch Previous [2, 29]	Pan Up [2, 6, 29, 44]	Maximize [6, 44]	Mute Speaker [2, 6, 36, 40]
Previous [2, 6, 29, 36, 40, 44]	Pan Down [2, 6, 29, 44]	Minimize [6, 44]	Unmute Speaker [2, 40]
Next [2, 6, 29, 36, 40, 44]	Zoom In [2, 6, 29, 44]	Answer Call [2, 29]	Volume up [6, 36, 40]
	Zoom Out [2, 6, 29, 44]	Hang up Call [2, 29]	Volume down [6, 36, 40]
			Play [6]
			Pause [6]
			Stop [6]
			Copy [2, 6, 44]
			Cut [2, 6, 44]
			Paste [2, 6, 44]
			Open Menu [6, 36, 40, 44]
			Close Menu [6, 36, 40, 44]

operations (*e.g.*, "*panning*" or "*zooming*") and non-visual-necessary tasks which users can perform without visual feedback (*e.g.*, "*volume down*" or "*turn on microphone*"). We followed the classification instructions in [2, 29], grouping the tasks into two categories: action and navigation. Within each category, there were two sub-categories: a task that can either be performed on a system/smartphone, computer or tablet (*e.g.*, switching to a different App) or a task that can be performed on a particular application system (*e.g.*, browsing websites). This classification allows us to create tasks that would be representative for head-worn devices such as smart earpieces or augmented glasses, while minimizing task duplication.

3.2 Participants

Twenty-eight paid participants (17 females) were recruited from our university and community volunteers for the study. The participants ranged in age from 20 to 34 years old (Mean = 24.75, SD = 2.93 years) and come from different academic backgrounds including engineering (N = 17), design (N = 5), humanities and social sciences (N = 5), and business (N = 1). Only one participant wear an earring on her earlobe. All participants are Asian, live in Taiwan, have owned or used smartphones and are familiar with touch and gesture input.

3.3 Procedure

At the beginning of the formal study, the experimenter explained the goals of this study and introduced the concepts of on-body and midair input proposed from related works to the participants via verbal description. This was done to ensure that all participants have a common understanding of ear-based input before eliciting the gestures.

Next, each participant was presented with a total of 31 tasks to execute. The order of tasks was random. Participants were given the goal of each task and an abstract animation and were then asked to design a gesture using the think-aloud protocol. The animations were provided to help the participants to better understand the function of each task. To avoid legacy bias [7], the animations were comprised of simple geometric shapes, sound, color and the task title which could allow participants to understand the meaning of tasks, including both visual-necessary and non-visual-necessary tasks. They instructed users by changing position, shape, or color, *e.g.*, participants can

Fig. 1. Our experiment was conducted within an (a) open environment using two laptops, and one of which showing the task's animation, *e.g.*, the animation of *"zoom out"* was displayed from (b) to (c) in one second.

have an idea of "*zoom out*" via objects getting smaller (see Figure 1 (b) and (c)) or "*close menu*" through the suddenly disappearing rectangle. This mitigates the effects of certain commonly-used GUI interfaces. Within the period of designing gestures for each task, the participants could view the animation as many times as they wanted, and were allowed to request clarification until he/she fully understood the function of the task.

Whenever the gesture for each task was defined, the participant received a list of exploratory questions about gesture details for better understanding their design rationales and preferences, such as "*Why is this gesture suitable?*" and "*Why did you choose this interaction type?*". Other questions were posed to the participants to follow up accordingly to their response. We also asked them to provide a continuous numeric rating from 1 to 7, with 1 indicating "strongly disagree" and 7 "strongly agree", according to the following criteria: (a) *Suitability*: the gesture I designed is a good match for its intended use. (b) *Usability*: The gesture I designed is easy to perform. (c) *Social Comfort*: I can perform this gesture in a social environment without feeling uncomfortable.

Note that participants could change their designs and track their previous gestures via video anytime if they thought of more suitable ones, which ensured design coherence. Only their final decisions were kept.

3.4 Study Configuration

The study configuration is displayed in Figure 1. Previous research [35] has shown that social environments affect participants' design rationales. Therefore, to make the resulting gesture set that is fitting for social scenarios, we conducted our experiment at a coffee shop or in a discussion room in our department. These locations are public to students and residents nearby, where everyone can talk and mingle freely. Note that our goal is to ensure that the environment is open and comfortable for the participants of the experiment. Thus, when the discussion room became too crowded and noisy, six participants moved to a coffee shop where they each agreed that both locations have similar levels of public exposure. The participants sat in front of two displays, where one showed the animation of a task, and the other showed the live recording of the experiment. Therefore, the participant could observe the gesture he/she created and modify it accordingly.

4 ANALYSES AND RESULTS

The study results include recorded videos, interview transcriptions, 31 tasks $\times$ 28 participants = 868 gestures, and the subjective ratings from the participants. Three researchers labeled these gestures independently. We took a screen shot of the gestures from the video record and attached to it the transcript of participants' feedback for each task, and then used an affinity diagram to analyze these gestures and discover any themes.

Table 2. Taxonomy of gestures for ear-based interaction gestures based on collected gestures and previous studies.

Gesture Mapping		
Nature	Metaphoric	Gesture is a metaphor of another object.
	Symbolic	Gesture visually depicts a symbol.
	Abstract	Gesture Mapping is arbitrary.
Context	Dependent	Gesture requires specific context.
	Independent	Gesture does not require specific context.
Flow	Continuous	Action occurs during the gesture.
	Discrete	Action occurs after the gesture completion.

Physical Characteristics		
Locale	Mid-air	Gesture occurs in the air with no physical contact.
	Touch-based	Gesture involves a contact with the ear.
	Mixed Locale	Gesture involves both locales.
Complexity	Simple	Gesture consists of a single gesture.
	Compound	Gestures can be decomposed into simple gestures.
Form	Static Pose	Hand pose is held in only one locale.
	Static Pose and Path	Hand pose is held as hand moves.
	Dynamic Pose	Hand pose changes in one location.
	Dynamic Pose and Path	Hand pose changes as hand moves.
	Deformation	Hand pose makes the ear deformation.

4.1 Taxonomy of the Gestures

The taxonomy was partially adopted from the previous studies [2, 26, 29], in which their taxonomy had two different classes of dimension: *gesture mapping* and *physical characteristics*. Gesture mapping describes how participants mapped the gestures to various tasks, including *nature*, *context*, and *flow*. Physical characteristic, on the other hand, captures the characteristics of gesture themselves, including *locale*, *complexity*, and *form*. The entire taxonomy is listed in Table 2. As for the inter-rater reliability, we further invited an independent rater who was shown the same categorization and did the classification for 372 gestures (12 trials were randomly picked for each task). The inter-rater reliability rate was calculated using Cohen's Kappa value (k = .889, p < .01) and was higher than .8, showing that it was sufficient to build the validity of our classification [15].

The *nature* reflects the different levels of semantic knowledge involved in the gestures. This dimension involves subcategories of *metaphoric*, *symbolic*, and *abstract* gesture types. Note that we did not use the *physical* gesture type, as no particular physical object was specified in this work, which is different from the smartwatches, mobile phones, and tables in previous studies [2, 29, 44]. The following are descriptions of the three categories:

- *Metaphoric*: The gesture acts on, with or like something else. In other words, it is a metaphor for another physical object. For instance, use a thumb to press a button on an imaginary stopwatch; spin the hand clockwise pretending to twist a knob to turn volume up.
- *Symbolic*: A symbolic gesture visually depicts a symbol, such as drawing a triangle in the air to execute "*play*."
- *Abstract*: Gesture mapping is arbitrary. It doesn't have any metaphorical or symbolic connection to the referent, *e.g.*, tapping on the ear with three fingers to stop the music.

The *context* dimension describes whether a gesture requires it being performed within a specific context, or being performed independently. For example, an index finger sliding down along the helix is context-specific. If you perform the gesture while listening to music, the volume will be lowered. By contrast, going back to the homes creen by tapping the earlobe twice is a context-independent gesture.

The *flow* dimension describes whether the gesture action on an object occurs simultaneously or after executing a gesture. The gesture would be categorized as a discrete gesture if the action occurs after the gesture is made, such as a tap on the ear to select the object. The gesture would be considered continuous if the action occurs while the gesture is on-going such as a swipe in the air to scroll the content.

The *locale* describes where, in relation to the human ear, a gesture is performed. The gesture is considered *touch-based* if it requires physical contact with the ear. In contrast to *touch-based* is *mid-air* if the gesture is performed in the air. A gesture that requires both is considered *mixed locale*.

The *complexity* dimension describes whether a gesture is a simple gesture or a compound gesture. A compound gesture can be decomposed into simple gestures by dividing spatial discontinuities. For instance, pinching the earlobe is a simple gesture, while pulling the earlobe is a compound gesture because it can be divided into two spatial discontinuous gestures (*i.e.*, pinching and pulling.) The scope of the *form* dimension is within one hand. In Piumsomboon's taxonomy [26], *form* contains only four categories; however, we add one category *Deformation* into *form* dimension because of the ear's specific affordance. The followings are descriptions of the five categories:

- *Static Pose*: The hand pose remains in only one locale, *e.g.*, putting a fist above the ear.
- *Static Pose and Path*: When the hand changes position, the posture of the hand remains the same (no finger movement), *e.g.*, moving the open palm away from the ear.
- *Dynamic Pose*: The gesture is defined as a dynamic pose when the hand's pose changes while the hand is still being held in the same position, *e.g.*, splaying the hand while keeping the hand above the ear.
- *Dynamic Pose and Path*: The hand changes its pose when the hand changes position, *e.g.*, splaying the hand while moving the hand away from the ear.
- *Deformation*: The gesture is classified as a deformation gesture when it makes the ear deformed, such as bending the ear or pulling the ear lobe.

Figure 2 shows the distribution of each dimension and illustrates the breakdown of our classifications of the 868 gestures.

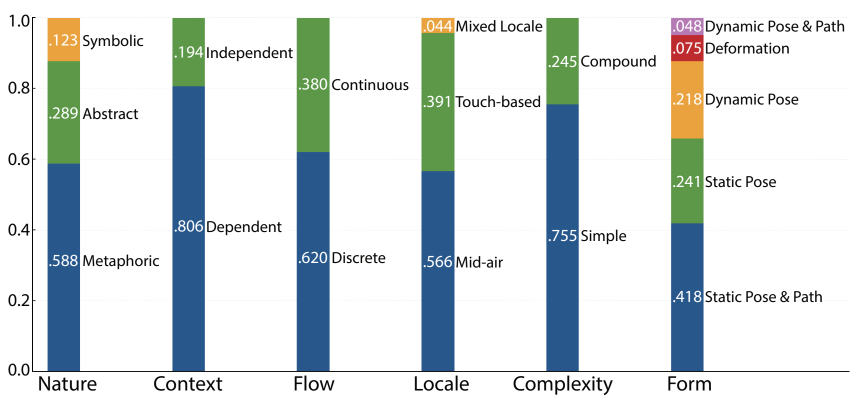

Fig. 2. The proportions of the 868 ear-based input gestures within each category in the six dimensional taxonomy; y-axis represents the percentage.

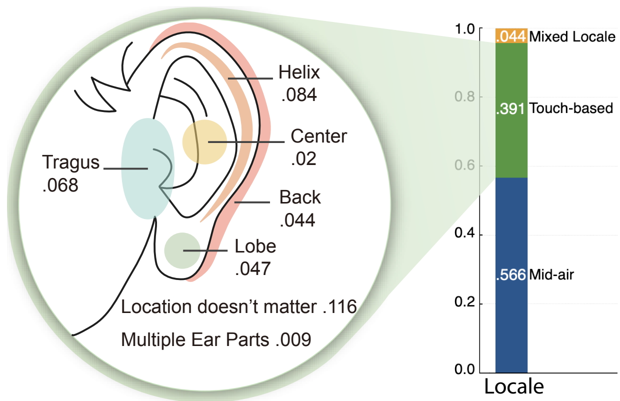

Fig. 3. The percentage distribution of different ear parts where participants had touched for each locale subcategory of this taxonomy.

To find out the preferences of ear parts in touch-based input, we labeled those gestures with the different parts of the ear where were touched, and there were finally 7 subcategories for touch-based input— *tragus*, *center*, *lobe*, *helix*, *back*, *multiple ear parts* and *location doesn't matter*. The definition of *lobe* and *helix* are the same as the physiological structure, but our *tragus* subcategory consists of the ear tragus as well as the junction of the tragus and cheek. Moreover, the *center* refers to the entire anterior outer ear except for the helix, lobe, and tragus. The gesture is considered *back* if it is in contact with any part of the posterior outer ear and the gesture that requires more than two of the above is considered *multiple ear parts*. The last special subcategory is called *location doesn't matter* which suggests that the participants did not care where they have touched but were focused on the action of "touch." The results (see Figure 3) suggest that the participants seemed to prefer completing the touch action rather than exactly touching a specific point.

4.2 User-defined Gestures

4.2.1 Classification Method. In the study of Wobbrock *et al.* [44], they group identical gestures for each task. And their consensus set of gestures consists of the most common gesture (the group consisting of largest number of any given gesture vis-à-vis others) elicited from participants for each task. Roughly following their approach, we first group the same gestures for each task. Next, we modified the classification according to whether you touch your ears. Later, when running an affinity diagram, we discovered two other factors that have an impact on the results – *gesture direction* and *number of fingers*, and then consider these two factors to regroup the gestures. After three grouping sessions, we obtained the final classification for each task. These two factors will be discussed below.

The gesture direction. Gesture direction represents the direction of movement relative to the ear. In the browsing tasks that are about switching, such as *next*, *previous*, *app switch next*, or *app switch previous*, many participants performed the same gesture but in different directions. After the semi-structured interviews, we found that the direction of gestures participants performed was influenced by their prior experience with the existing user interface. For example, when designing a gesture for *app switch next*, some participants swiped along the x-axis and some swiped along the y-axis. The direction of the gestures corresponded to their previous experiences relative to transition animation as shown on a touchscreen [2]. Although we have tried to avoid legacy bias by removing elements that were related to mobile phones and computers, it was still difficult to prevent participants from being influenced by their previous experience. Therefore, we decided not to consider which axis but to consider only the positive or negative of the axis that the gesture moves along.

The number of fingers. The number of fingers represents the fingers that the participant considered related to the gesture. This is because we found that when participants generated gestures by using a different number of fingers for similar tasks, they often only cared if the number of fingers used one, or more than one, instead of a precise number of fingers. One participant even said that he knew he used more than one finger but did not remember the number accurately because that did not matter to him. This observation is similar to previous studies wherein participants had little concern about how many fingers were used in a gesture [6, 45]. To cope with this confusion, we simply separated gestures that used two or more fingers from those that used only one finger. This is less constraining than matching the exact number of fingers used in performing gestures. But there was one exception, when the gesture was a metaphor (*e.g.*, using two fingers to perform the scissor), not abstractly used, we would not follow the aforementioned criteria. Loosening the restriction from "gestures must be identical within each group" to "gestures must be similar within each group" made this classification better represent the thought underlying the gestures.

4.2.2 Agreement Between Participants. For each task, we placed groups of similar gestures together. The group size was then used to compute an agreement score which can display the level of participants' consensus. We adopted the revised version of an agreement rating formula from Vatavu and Wobbrok *et al.* [39] which accounts for a degree of freedom and made the agreement rate more representative given the large samples of participants with the same proposal ratios. The revised version of the formula is defined as the follows:

$$AR\left(r\right) = \frac{|P|}{|P|-1}\sum_{P_i \subseteq P}\left(\frac{|P_i|}{|P|}\right)^2 - \frac{1}{|P|-1} \tag{1}$$

In Eq. (1), P is the set of all proposals for task r, $|P|$ is the total number of gesture within the task r, and $|Pi|$ is a subset of similar proposals from P, and the range of AR values in [0,1].

Figure 4 illustrates the agreement rates for all 31 tasks. According to the qualitative interpretations for agreement rates proposed by Vatavu *et al.* [39], our agreement rates ranged from .093 (low agreement, $AR \leq .1$) to .526 (very high agreement, $AR > .5$). The mean of AR was .213 (medium agreement, $.1 < AR < .3$).

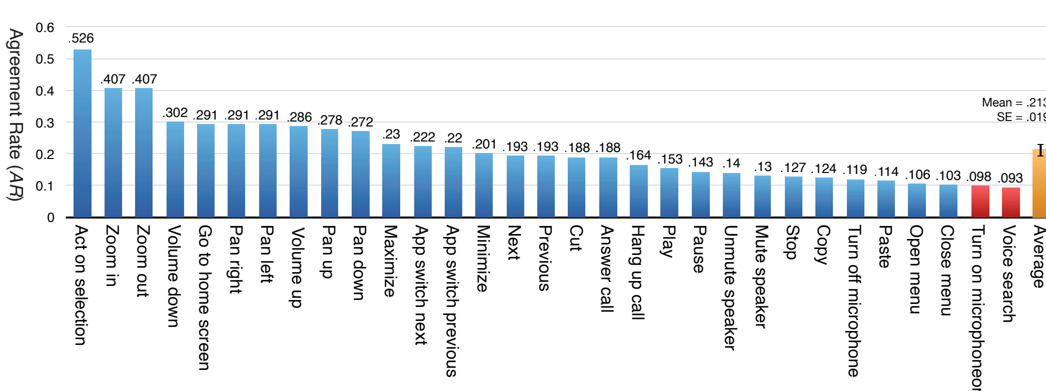

Fig. 4. Agreement rates for the 31 tasks in descending order.

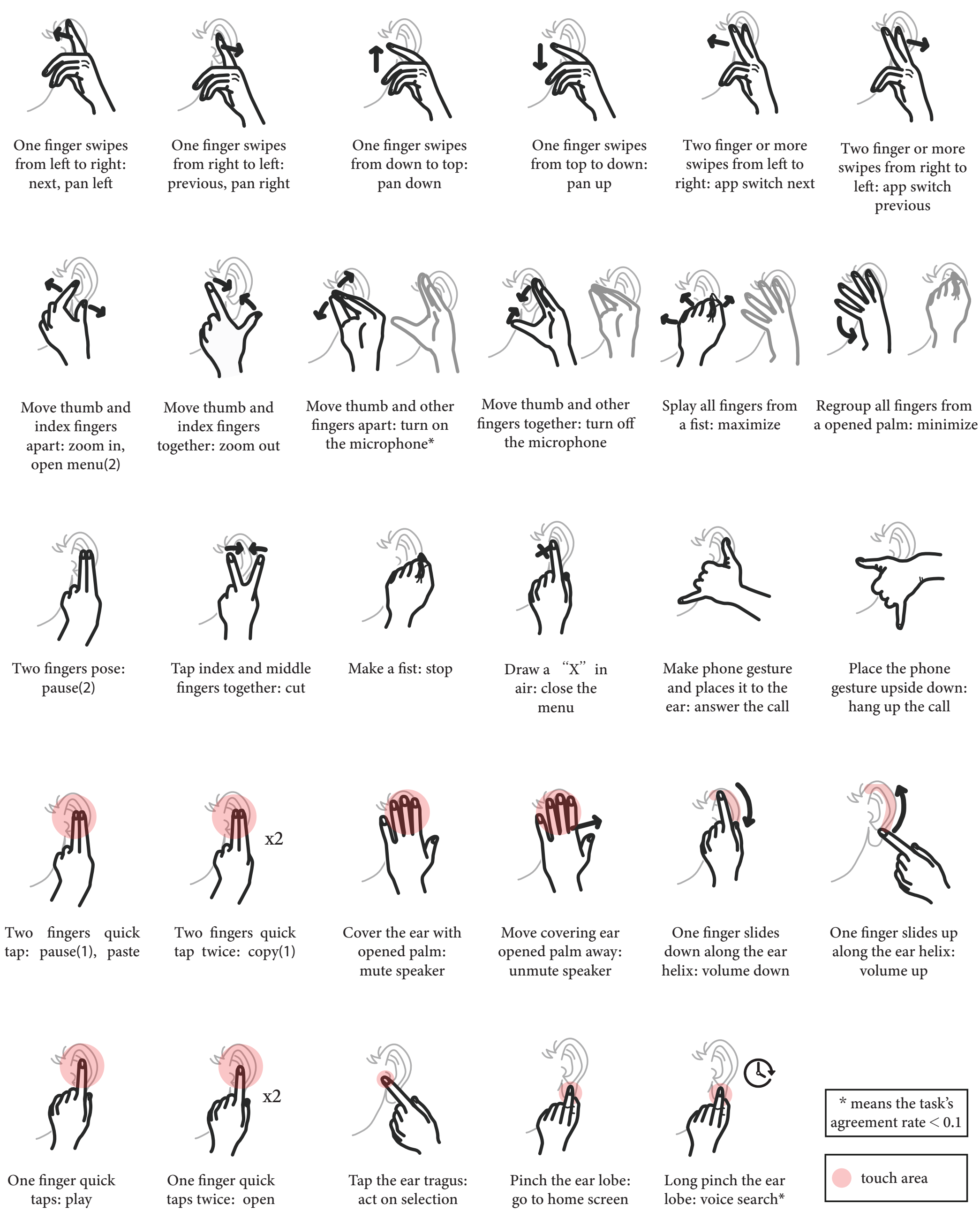

Fig. 5. Our study's user-defined gesture set for ear-based input. The touch area indicates the command trigger area, *e.g.*, two fingers quickly tap anywhere on the ear to trigger "*pause*" or one finger tap the ear tragus for an "*act on selection*" task. Note that all gesture representations meet the user-driven design principles which were presented by McAweeney *et al.* [21].

4.2.3 A User-defined Gesture Set. After grouping similar gestures in accordance with the aforementioned criteria, the gesture found among the biggest group for each task is considered to be the *representative gesture*. We referred to the set of those *representative gestures* as our *user-defined gesture set*. Naturally, the gestures that were not found among the biggest group of each task were automatically classified as the *discarded set*. Note that since we set two special criteria as mentioned herein prior, the biggest group of some tasks included a gesture with different directions or numbers of fingers. In such cases, we chose that particular gesture that was found in a relative majority from the group set among the biggest group assigning it as the *representative gesture*.

From this user-defined gesture set, we found that there were three tasks which were assigned two gestures to them ('*pause*", "*copy*" and "*open menu*"), and 28 tasks only had one gesture. The reason these three tasks were assigned two gestures was that there were two groups with the same and equally large consensus in gestures for those tasks.

The complete user-defined gesture set is illustrated in Figure 5. This gesture set is a non-one-to-one mapping of tasks and gestures. To illustrate precisely the gestures, we followed the design principles for gesture representations which were presented by McAweeney *et al.* [21]. In addition, despite there being no consensus for "*voice search*" (AR = .093) or "*turn on the microphone*" (AR = .098), we still provided the gesture which comprised a relative majority and marked these with a star to distinguish this quality.

4.3 The Subjective Rating of Gestures

The average scores for the *user-defined gesture set* and *discarded set* respectively were 5.85 (σ = .37) and 5.75 (σ = .28) for *Suitability*, 5.91 (σ = .36) and 5.86 (σ = .25) for *Usability*, 5.57 (σ = .49) and 5.55 (σ = .38) for *Social Comfort*. Comparing the subject ratings, the subjective scores for *Suitability* of the *user-defined gesture set* was significantly higher than the *discarded set* (Z = 2.949, p = .003). However, there was no significant difference for *Usability* and *Social Comfort* between both groups. The percentage and score distribution of subject ratings were displayed in Figure 6. In general, the participants consider that their gestures are suitable and plausible to be performed within social scenarios. We speculate that due to the study environment of this study, the *Usability* and *Social Comfort* for the consensus set and the discarded set are close. In other words, most participants considered social-comfort and ease when designing their gestures in a semi-public environment.

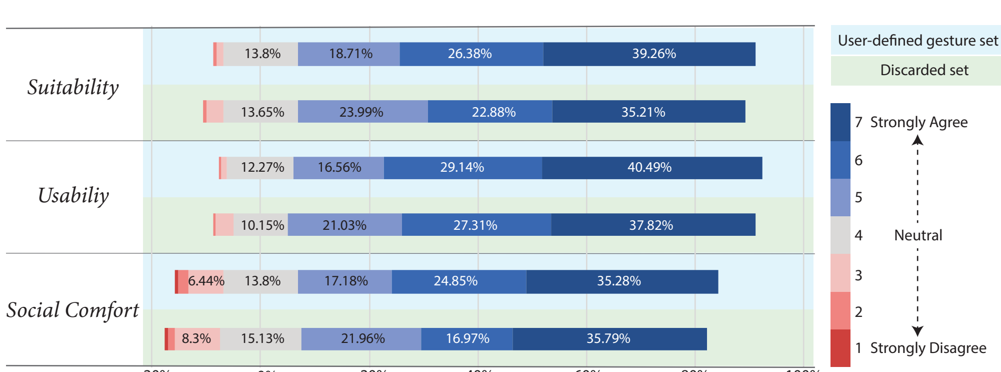

Fig. 6. The percentage and score distribution of subject ratings for the user-defined gesture set (blue background) and the discarded set (green background); x-axis represents the percentage. In addition, the exact percentage less than 5% is not shown.

5 DESIGN RATIONALES, PREFERENCES, AND DISCUSSIONS

After analyzing the data including transcripts and video recording, we identified the following common themes which describe the design rationales and preferences of the participants and notable findings.

5.1 Migrating Touch Screen Gestures

Familiar touch gestures are often considered. For example, some participants swiped in the air for panning and moving thumb and index finger apart or together for zooming as if they were using a touchscreen. In addition, for some input actions such as *"play," "act on selection"* and *"go to home screen,"* they tapped with a fingertip as if there were an imaginary surface next to or on the ear. Mimicking of touchscreen gestures was also referred to in previous studies [2, 35]. Participants often believed that these touchscreen related gestures were a good match for the tasks. From the interview, the participants reported that *"imitating the touchscreen gestures is easier to perform and easier to remember (P17, P19-21, P25)"* because these gestures have been standardized and used for a long time. Finally, there was a high level of agreement rates on those gestures which mimic touchscreen gestures (See Figure 4 and Figure 5.)

5.2 Utilizing Sign Gestures with Real-world Metaphors

As previous studies have referred [6, 26, 29, 35], participants also incorporated sign gestures in their design. For example, to "*answer a call*," a majority of participants performed a phone gesture which was made of bending the index, middle and ring fingers and place the hand in this pose close to the ear. By contrast, for "*hang up a call*," participants removed the "*answer a call*" gesture from the ear or turned the phone gesture upside down. The participants described that it is more intuitive to imitate the conventional telephone experience. The other example is that participants performed an opened palm for *stop* because this gesture was intuitive when you wanted to stop someone from approaching you, and it is also used for stopping cars when directing the traffic. These metaphorical gestures and sign gestures were usually considered a better fit for the task. The participants explained that metaphorical gestures were more concrete and recognizable so that they could also inform bystanders what they were doing, as other gestures may decrease their social comfort.

5.3 Preference of Interaction Methods for Different Tasks

Since we did not have any constraint on the interaction type in this user elicitation study, participants intuitively designed different interaction methods and gestures for ear-based input. As a result, the gestures that participants designed included both midair and touching the ear gestures. Some participants performed gestures involving touching the ear because its *"tactile feedback made them feel the recognition more accurately (P10, P12-13, P16-17, P20, P28)"*, and *"touching the ear is a natural behavior without attracting other people's attention (P5, P10, P21)."* By contrast, others commented that *"performing gestures in the air is better due to its spatial freedom to display gesture variety (P8, P14, P18)."*

The data in shown in Figure 2 suggests that regarding ear-based input, touch-based and midair gestures are equally important (43.4% vs. 56.6%) regarding the number of gestures involving the ear (touch-based and mixed locale) and gestures only in midair. Nonetheless, we found that 75% of participants designed midair gestures when it comes to the tasks about browsing or adjusting scale, such as *panning*, *zooming in*, *zooming out*, *maximizing* and *minimizing*. P13, P16, and P17 said that *"I did a midair gesture if the tasks used frequently. If I always perform it on my ear, it would be a little uncomfortable."* On the contrary, an average of 70% of participants tended to design touch-based gestures for some tasks regarding action triggering, *i.e.*, *act on selection* and *go to home screen*. In addition, an average of 68% of participants performed gestures involving touching the ear for tasks that adjust sound, *i.e.*, *volume up/down* and *mute/unmute speaker*. For instance, participants who designed the gesture, covering the ear by palm, for *mute speaker* said that *"Covering the ears is a natural act when you hear a loud sound (P2, P4-5, P14, P17, P24, P28)."* Interestingly, participants adjusted the speaker through the medium—the ear, the organ of hearing in mammals.

In brief, as per the feedback from participants, instead of performing all task gestures in midair or in a touch-based way, they showed more willingness to performing gestures in a different way according to the tasks' type.

5.4 Design rationales of Touched-based Gestures

Two kinds of design rationales were usually seen when participants designed touch-based gestures. One is performing a metaphoric or symbolic gesture to touch the ear for making sure it was completely performed or, they imagined *"touching would be detected more accurately (P10, P11-12, P22)."*

The other design rationale was using the point or locale where gesture is performed to assign tasks. This kind of design rationale was usually seen when participants were trying to transfer the smart device experience to ear-based input. To further explore if this design rationale regarding which points on the ear were used frequently and whether the ear points were mapped for specific tasks would reach a consensus among participants, we analyzed the gestures and found that the *lobe* (17.5%), *tragus* (25.9%), *helix* (32.0%), and *back* (16.7%) are the most preferred areas to perform touch-based gestures upon with a specified touch point in descending order of preference, and their use corresponded to specific tasks. For example, in proposed helix-based gestures, 52.1% were performed with a sliding gesture along the helix such as adjustment type tasks, adjusting "*volume up*" and "*volume down*." A majority of participants pinched the ear lobe to "*go to home screen*" because they considered that the ear lobe is special and obvious, and its round shape made them think of the "home button." In addition, the ear tragus was usually chosen for "*act on selection*," or other tasks for which participants considered that the task could be actuated by pressing a button.

5.5 Similar Gestures for Seemingly Related Tasks

As previous studies have shown [6, 26], we had a similar finding in that participants chose similar gestures for seemingly related tasks or the tasks that they thought it would be executed consecutively. They tried to set rules for performing the gestures of related tasks. One pattern was to perform the same pose but use a different interaction types such as tapping the index and middle finger together for actuating "*cut*" and tapping the ear with these two fingers for actuating "*paste*." Another pattern was to perform the same interaction types but to do so by using a varying numbers of fingers, *e.g.*, some participants used one finger to swipe right for *next* and used two or more fingers to swipe right for *app switch next*. It makes the series of tasks more convenient and faster to perform and easier to remember if gestures exist using the same rules or same forms. Furthermore, one participant proposed a special pattern that mapped the number of fingers to the order that in which tasks were generally executed. More specifically, this participant designed gestures with finger increment patterns, assigning one, two fingers and five finger tapping for the "*play*," "*pause*" and "*stop*" task because the button was usually pressed in sequence.

Interestingly, for those tasks which were different, but participants felt had similar attributes, they were also assigned similar gestures. In our user-defined gesture set, swiping is preferred for the tasks *"pan," "next," "app switch next," and "volume up,"* and splaying the hand is preferred for the tasks *"zoom in," "maximize," and "turn on microphone."* Besides, some participants mentioned

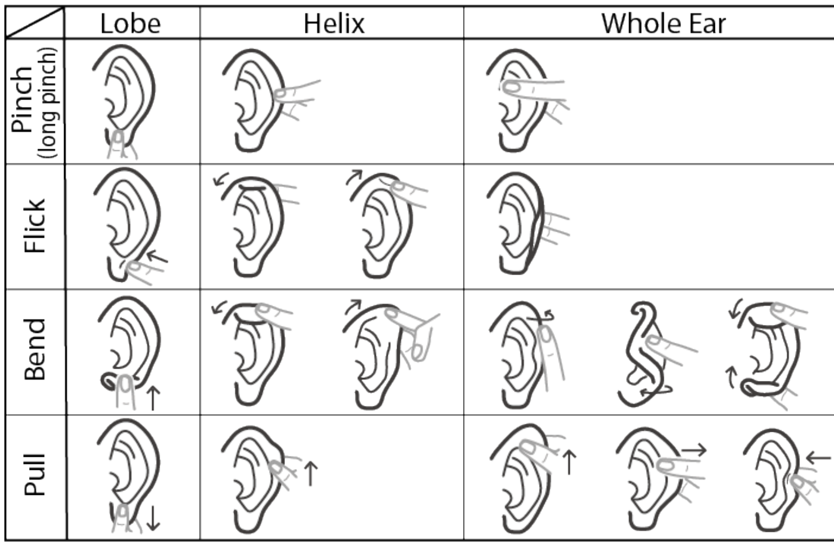

Fig. 7. Design space for deformation interaction for ear-based input, organized primarily by where (columns) and how (rows) the deformation takes place and the locale at which the corresponding method may be performed.

that the gestures should be the same when performing "*on/off*" and "*mute/unmute,*" like clicking a switch.

5.6 Design Space of Deformation Interaction

Because the outer ear is made of ridged cartilage covered by skin, it is flexible enough that some participants deformed the ear for ear-based input. This form was chosen for 7.5% of all the 868 gestures. Only two tasks, "*go to home screen*" and "*voice search,*" using such gestures are among our user-defined gesture set. We believe that the deformation gestures are used "sparingly" to avoid physical and social discomfort. Nevertheless, we are also interested in how participants deformed the ear and at which points did the participants deform their ear, so we classified all deformation gestures among the 868 gestures and then provided a design space. We found that if we only considered the deformation type when analyzing gestures, there were four methods of deforming the ear—pinching, flicking, bending and pulling. Those actions were also varied by duration, deformation quantity, and direction. Moreover, participants tended to deform specific ear parts, namely the lobe or the helix, or alternatively the entire ear. Figure 7 shows the design space in regard to deformation interaction for ear-based input.

5.7 Preferences regarding Fingers for Performing Gesture

From all 868 gestures, we found that a majority of gestures consist of using a single finger (39.5%), while using two fingers (31.6%) or the palm (22.6%) comprised the bulk of the remaining gestures. In terms of single-finger gestures, participants chose to use the index finger (90.6%), the thumb (8.7%), and the middle finger (0.6%) in that descending order of preference while no participant used their ring finger or the picky finger independently. When it comes to two-finger gestures, there are four finger pairs in use— thumb and index finger (50.7%), index and middle finger (41.6%), thumb and pinky (7.3%), and index finger and pinky (0.4%). And the most frequently used fingers are the thumb (47.1%), index finger (91.4%) and middle finger (39.5%) in that order. This could be explained by Wolf *et al.*'s [46] assertion in their summary of the anatomy of the hand it is that because of biomechanics, that the thumb, index finger and middle finger are used more as they are more dexterous and suitable for independent movement. Alternatively, the ring finger is considered clumsy because there are two muscles synergistically bending the index, middle, and little finger to bring them into the palm position. In addition, social connotations of given gestures also had an impact on which fingers participants choose to use. For example, one participant expressed that he avoided performing gestures with a single middle finger because of its insulting meaning. Neither did he use the pinky finger. Although the pinky finger can move independently, it was less used separately due to an assigned offensive social meaning by the users.

5.8 Comfort to perform gestures in public

As mentioned prior, sign gestures and metaphorical gestures also were to be less comfortable to use in public because their meaning was known to all. Moreover, two participants indicated that the gestures that are multi-stepped or taking a long time to perform were generally not acceptable to participants, as well. However, those gestures that were seemed to slightly touch the ear such as pinching the ear lobe or touching the ear helix and sliding down as if tucking hair behind the ear were considered to be more suitable to perform in public than midair gestures. Also, nine participants indicated that the small-sized gestures that only involve wrist movement, not the entire arm, and with the performing range not exceeding the width of their shoulder to be more acceptable.

5.9 Comparison to Previously Proposed Ear Gestures

In the condition where users have a larger input area [33], ears are used for ignoring a call task. Serrano *et al.* [33] also show that legacy gestures are often considered for general tasks, such as panning, zooming, and rotating. This echo to our user study results. For more future work, it would be interesting to re-examine how face-based and ear-based gestures can be modified and combined into a more powerful yet intuitive gesture set.

5.10 Implications for Gesture Recognition of Ear-based Input

Similar to other user elicitation studies, we focused on exploring user-defined gestures, regardless of the sensory technique. Although we did not consider how to track gestures and where to place the sensing device on, we still offer several suggestions concerning the technique that is required to recognize the gestures from the results of this work. To be more specific, our user-defined gesture set indicates that people utilize the midair method to perform 58% of tasks as mentioned in Table 1. For midair input, the gestures include static and dynamic poses, and the number of fingers plays a vital role. Therefore, the technique of hand and finger tracking and recognition are required. Soli [17] is a possible solution for the midair gestures in our work. Also, depending on the users' design preferences, the sensing region that is directly above the shoulder and below the head is required at least. In comparison with the gestures made in the air, users consider the surface of the outer ear as an input device during performing touch-based gestures. Hence, we recommend that the sensing area should cover the entire ear flexibly and enable the sensors to detect the position and number of touchpoints. On-body electronic signal sensing, ActiTouch [48], is a viable solution for touch-based gestures. In addition, some prior works can implement a sensing method for those deformable gestures in this work, *e.g.*, both EarPut [18] and EarTouch [14] detect large and small deformations.

6 LIMITATIONS AND FUTURE WORK

As previous studies indicate [34, 40], legacy bias may have an impact on the results of this user elicitation study. Although we took care to not show any smartphone or computer related elements within our task animations to offset the legacy bias from the touchscreen, participants still often utilized the smart device paradigm. It could be seen that touchscreen gestures were usually referred to mentally by the participants and led to a greater agreement amongst them. Additionally, the legacy bias from personal experience impacts the agreement scores. There was an example that previous experience negatively affected agreement scores: the gestures that participants designed for "*pause*" included using two fingers tapping to symbolized the *pause* icon, showing an opened palm forward as in traffic sign language and pressing an imaginary pause button. These gestures showed how participants transferred their personal experiences into gesture design. This observation also suggests that the user-defined gesture may vary according to the background of the participants, in regard to such factors as age, culture, and gender. Examining these factors is important future work.

Also, there is another limitation of this work that this study was only conducted with the participants sitting. With the popularity of smart headphones, people often wear them in a variety of environments (such as running.) Besides, a previous work [4] showed that compared to running, participants thought that the same on-body input gestures were easier to perform when standing. Therefore, more exploration is needed to understand how the position of such participants affects ear-based input.

The other limitation of our study is that we did not take the personal difference into account, such as wearing an earring. One participant mentioned that because of her earrings, when she considers the earlobe to be a good place for gestures, she moves around to the edge of the earlobe rather than the center of the earlobe. Hence, the personal differences of the ear should be further studied in the future.

7 CONCLUSION

Our study explored user-defined gesture input for smart earpieces beyond the capabilities of recent sensing technology. We proposed a user-defined gesture set based on participants' agreement in regard to 868 gestures and summarized the interaction methods that participants designed in our user elicitation study. Our results indicate that participants liked both midair gestures and touch-based gestures, and that they chose different interactions depending on the type of tasks. Furthermore, we discovered participants' design rationales, preferences, and gained insight into design patterns which could be translated into implications for sensing technology development and interaction design. We believe that our study represents a necessary step towards making ear-based input devices more intuitive and effective.

ACKNOWLEDGMENTS

All this research is done and supported by National Chiao Tung University and National Taiwan University. We thank our reviewers for the helpful feedback.

REFERENCES

[1] Toshiyuki Ando, Yuki Kubo, Buntarou Shizuki, and Shin Takahashi. 2017. CanalSense: Face-Related Movement Recognition System Based on Sensing Air Pressure in Ear Canals. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. ACM, New York, NY, USA, 679–689. https://doi.org/10.1145/3126594.3126649

[2] Shaikh Shawon Arefin Shimon, Courtney Lutton, Zichun Xu, Sarah Morrison-Smith, Christina Boucher, and Jaime Ruiz. 2016. Exploring Non-touchscreen Gestures for Smartwatches. In *Proceedings of the 2016 CHI Conference on*

Human Factors in Computing Systems (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 3822–3833. https://doi.org/10.1145/2858036.2858385

[3] Daniel Ashbrook, Carlos Tejada, Dhwanit Mehta, Anthony Jiminez, Goudam Muralitharam, Sangeeta Gajendra, and Ross Tallents. 2016. Bitey: An Exploration of Tooth Click Gestures for Hands-Free User Interface Control. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 158–169. https://doi.org/10.1145/2935334.2935389

[4] Roger Boldu, Alexandru Dancu, Denys J.C. Matthies, Pablo Gallego Cascón, Shanaka Ransir, and Suranga Nanayakkara. 2018. Thumb-In-Motion: Evaluating Thumb-to-Ring Microgestures for Athletic Activity. In *Proceedings of the Symposium on Spatial User Interaction* (Berlin, Germany) *(SUI '18)*. Association for Computing Machinery, New York, NY, USA, 150–157. https://doi.org/10.1145/3267782.3267796

[5] Idil Bostan, Oğuz Turan Buruk, Mert Canat, Mustafa Ozan Tezcan, Celalettin Yurdakul, Tilbe Göksun, and Oğuzhan Özcan. 2017. Hands As a Controller: User Preferences for Hand Specific On-Skin Gestures. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh, United Kingdom) *(DIS '17)*. ACM, New York, NY, USA, 1123–1134. https://doi.org/10.1145/3064663.3064766

[6] Edwin Chan, Teddy Seyed, Wolfgang Stuerzlinger, Xing-Dong Yang, and Frank Maurer. 2016. User Elicitation on Single-hand Microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 3403–3414. https://doi.org/10.1145/2858036.2858589

[7] Julien Epps, Serge Lichman, and Mike Wu. 2006. A Study of Hand Shape Use in Tabletop Gesture Interaction. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI EA '06)*. ACM, New York, NY, USA, 748–753. https://doi.org/10.1145/1125451.1125601

[8] Chris Harrison, Hrvoje Benko, and Andrew D. Wilson. 2011. OmniTouch: Wearable Multitouch Interaction Everywhere. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) *(UIST '11)*. ACM, New York, NY, USA, 441–450. https://doi.org/10.1145/2047196.2047255

[9] Da-Yuan Huang, Liwei Chan, Shuo Yang, Fan Wang, Rong-Hao Liang, De-Nian Yang, Yi-Ping Hung, and Bing-Yu Chen. 2016. DigitSpace: Designing Thumb-to-Fingers Touch Interfaces for One-Handed and Eyes-Free Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 1526–1537. https://doi.org/10.1145/2858036.2858483

[10] Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. 1985. Direct Manipulation Interfaces. *Hum.-Comput. Interact.* 1, 4 (Dec. 1985), 311–338. https://doi.org/10.1207/s15327051hci0104_2

[11] 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* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 3015–3018. https://doi.org/10.1145/2702123.2702572

[12] Maria Karam and m. c. schraefel. 2005. A Taxonomy of Gestures in Human Computer Interactions. (2005). https://eprints.soton.ac.uk/261149/

[13] Frederic Kerber, Markus Löchtefeld, Antonio Krüger, Jess McIntosh, Charlie McNeill, and Mike Fraser. 2016. Understanding Same-Side Interactions with Wrist-Worn Devices. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (Gothenburg, Sweden) *(NordiCHI '16)*. Association for Computing Machinery, New York, NY, USA, Article 28, 10 pages. https://doi.org/10.1145/2971485.2971519

[14] Takashi Kikuchi, Yuta Sugiura, Katsutoshi Masai, Maki Sugimoto, and Bruce H. Thomas. 2017. EarTouch: Turning the Ear into an Input Surface. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. ACM, New York, NY, USA, Article 27, 6 pages. https://doi.org/10.1145/3098279.3098538

[15] Mary L McHugh. 2012. Interrater reliability: The kappa statistic. *Biochemia medica : časopis Hrvatskoga društva medicinskih biokemičara / HDMB* 22 (10 2012), 276–82. https://doi.org/10.11613/BM.2012.031

[16] DoYoung Lee, Youryang Lee, Yonghwan Shin, and Ian Oakley. 2018. Designing Socially Acceptable Hand-to-Face Input. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. ACM, New York, NY, USA, 711–723. https://doi.org/10.1145/3242587.3242642

[17] Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: Ubiquitous Gesture Sensing with Millimeter Wave Radar. *ACM Trans. Graph.* 35, 4, Article 142 (July 2016), 19 pages. https://doi.org/10.1145/2897824.2925953

[18] Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, Suranga Nanayakkara, and Max Mühlhäuser. 2014. EarPut: Augmenting Ear-worn Devices for Ear-based Interaction. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (Sydney, New South Wales, Australia) *(OzCHI '14)*. ACM, New York, NY, USA, 300–307. https://doi.org/10.1145/2686612.2686655

[19] Denys J. C. Matthies, Simon T. Perrault, Bodo Urban, and Shengdong Zhao. 2015. Botential: Localizing On-Body Gestures by Measuring Electrical Signatures on the Human Skin. In *Proceedings of the 17th International Conference on*

Human-Computer Interaction with Mobile Devices and Services (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 207–216. https://doi.org/10.1145/2785830.2785859

[20] Denys J. C. Matthies, Bernhard A. Strecker, and Bodo Urban. 2017. EarFieldSensing: A Novel In-Ear Electric Field Sensing to Enrich Wearable Gesture Input through Facial Expressions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1911–1922. https://doi.org/10.1145/3025453.3025692

[21] Erin McAweeney, Haihua Zhang, and Michael Nebeling. 2018. User-Driven Design Principles for Gesture Representations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 547, 13 pages. https://doi.org/10.1145/3173574.3174121

[22] C. Metzger, M. Anderson, and T. Starner. 2004. FreeDigiter: a contact-free device for gesture control. In *Eighth International Symposium on Wearable Computers*, Vol. 1. 18–21. https://doi.org/10.1109/ISWC.2004.23

[23] Meredith Ringel Morris, Jacob O. Wobbrock, and Andrew D. Wilson. 2010. Understanding Users' Preferences for Surface Gestures. In *Proceedings of Graphics Interface 2010* (Ottawa, Ontario, Canada) *(GI '10)*. Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 261–268. http://dl.acm.org/citation.cfm?id=1839214.1839260

[24] Miguel A. Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. 2013. Memorability of Pre-designed and User-defined Gesture Sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 1099–1108. https://doi.org/10.1145/2470654.2466142

[25] Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum. 2004. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In *Gesture-Based Communication in Human-Computer Interaction*, Antonio Camurri and Gualtiero Volpe (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 409–420.

[26] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-defined Gestures for Augmented Reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. ACM, New York, NY, USA, 955–960. https://doi.org/10.1145/2468356.2468527

[27] 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

[28] 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* (Atlanta, Georgia, USA) *(CHI '10)*. ACM, New York, NY, USA, 887–896. https://doi.org/10.1145/1753326.1753458

[29] 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* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 197–206. https://doi.org/10.1145/1978942.1978971

[30] T. Scott Saponas, Desney S. Tan, Dan Morris, Ravin Balakrishnan, Jim Turner, and James A. Landay. 2009. Enabling Always-available Input with Muscle-computer Interfaces. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology* (Victoria, BC, Canada) *(UIST '09)*. ACM, New York, NY, USA, 167–176. https://doi.org/10.1145/1622176.1622208

[31] Douglas Schuler and Aki Namioka (Eds.). 1993. *Participatory Design: Principles and Practices.* L. Erlbaum Associates Inc., Hillsdale, NJ, USA.

[32] Julia Schwarz, Chris Harrison, Scott Hudson, and Jennifer Mankoff. 2010. Cord Input: An Intuitive, High-accuracy, Multi-degree-of-freedom Input Method for Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. ACM, New York, NY, USA, 1657–1660. https://doi.org/10.1145/1753326.1753573

[33] 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* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 3181–3190. https://doi.org/10.1145/2556288.2556984

[34] Teddy Seyed, Chris Burns, Mario Costa Sousa, Frank Maurer, and Anthony Tang. 2012. Eliciting Usable Gestures for Multi-display Environments. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces* (Cambridge, Massachusetts, USA) *(ITS '12)*. ACM, New York, NY, USA, 41–50. https://doi.org/10.1145/2396636.2396643

[35] Ying-Chao Tung, Chun-Yen Hsu, Han-Yu Wang, Silvia Chyou, Jhe-Wei Lin, Pei-Jung Wu, Andries Valstar, and Mike Y. Chen. 2015. User-Defined Game Input for Smart Glasses in Public Space. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 3327–3336. https://doi.org/10.1145/2702123.2702214

[36] Radu-Daniel Vatavu. 2012. User-defined Gestures for Free-hand TV Control. In *Proceedings of the 10th European Conference on Interactive TV and Video* (Berlin, Germany) *(EuroITV '12)*. ACM, New York, NY, USA, 45–48. https://doi.org/10.1145/2325616.2325626

[37] Radu-Daniel Vatavu. 2013. A Comparative Study of User-Defined Handheld vs. Freehand Gestures for Home Entertainment Environments. *J. Ambient Intell. Smart Environ.* 5, 2 (March 2013), 187–211.

[38] Radu-Daniel Vatavu. 2019. The Dissimilarity-Consensus Approach to Agreement Analysis in Gesture Elicitation Studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300454

[39] 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* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 1325–1334. https://doi.org/10.1145/2702123.2702223

[40] Radu-Daniel Vatavu and Ionut-Alexandru Zaiti. 2014. Leap Gestures for TV: Insights from an Elicitation Study. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (Newcastle Upon Tyne, United Kingdom) *(TVX '14)*. ACM, New York, NY, USA, 131–138. https://doi.org/10.1145/2602299.2602316

[41] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O. Wobbrock. 2020. A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 855–872. https://doi.org/10.1145/3357236.3395511

[42] 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* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 179–188. https://doi.org/10.1145/2556288.2557239

[43] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the Guessability of Symbolic Input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. ACM, New York, NY, USA, 1869–1872. https://doi.org/10.1145/1056808.1057043

[44] 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* (Boston, MA, USA) *(CHI '09)*. ACM, New York, NY, USA, 1083–1092. https://doi.org/10.1145/1518701.1518866

[45] Katrin Wolf, Anja Naumann, Michael Rohs, and Jörg Müller. 2011. Taxonomy of Microinteractions: Defining Microgestures Based on Ergonomic and Scenario-Dependent Requirements. In *Proceedings of the 13th IFIP TC 13 International Conference on Human-Computer Interaction - Volume Part I* (Lisbon, Portugal) *(INTERACT'11)*. Springer-Verlag, Berlin, Heidelberg, 559–575.

[46] Katrin Wolf, Anja Naumann, Michael Rohs, and Jörg Müller. 2011. Taxonomy of Microinteractions: Defining Microgestures Based on Ergonomic and Scenario-Dependent Requirements. In *Proceedings of the 13th IFIP TC 13 International Conference on Human-Computer Interaction - Volume Part I* (Lisbon, Portugal) *(INTERACT'11)*. Springer-Verlag, Berlin, Heidelberg, 559–575.

[47] Xing-Dong Yang, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2012. Magic Finger: Always-available Input Through Finger Instrumentation. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (Cambridge, Massachusetts, USA) *(UIST '12)*. ACM, New York, NY, USA, 147–156. https://doi.org/10.1145/2380116.2380137

[48] Yang Zhang, Wolf Kienzle, Yanjun Ma, Shiu S. Ng, Hrvoje Benko, and Chris Harrison. 2019. ActiTouch: Robust Touch Detection for On-Skin AR/VR Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1151–1159. https://doi.org/10.1145/3332165.3347869

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

[50] Kening Zhu, Xiaojuan Ma, Haoyuan Chen, and Miaoyin Liang. 2017. Tripartite Effects: Exploring Users' Mental Model of Mobile Gestures under the Influence of Operation, Handheld Posture, and Interaction Space. *International Journal of Human–Computer Interaction* 33, 6 (2017), 443–459. https://doi.org/10.1080/10447318.2016.1275432

Received July 2020; revised August 2020; accepted September 2020