Seeing with the Hands: A Sensory Substitution That Supports Manual Interactions

Shan-Yuan Teng* Gene S-H Kim* Xuanyou Liu* Pedro Lopes

University of Chicago (*equal contribution)

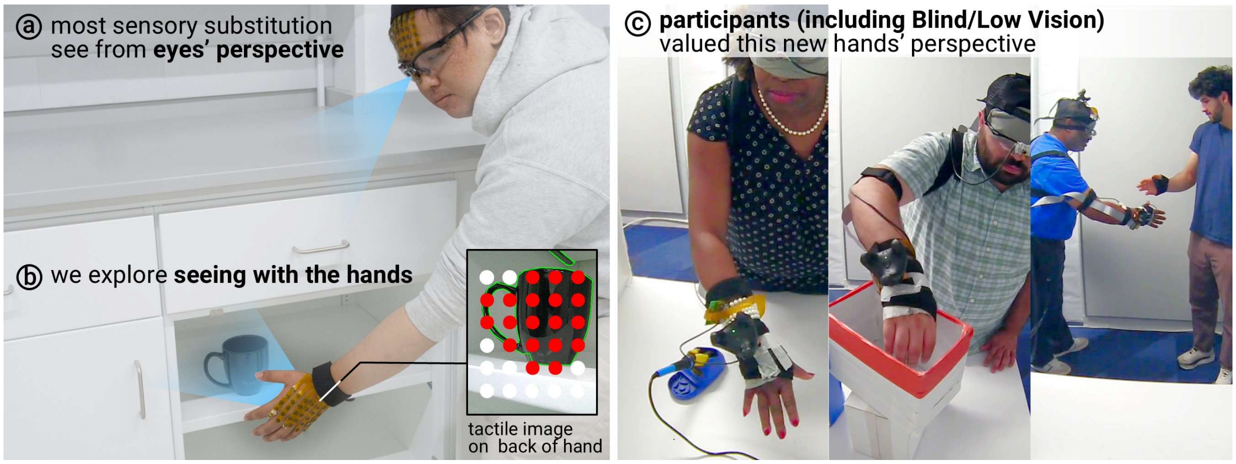

Figure 1: (a) Many sensory substitutions allow users to see from the eyes' perspective—camera & tactile array on the forehead. To understand the value of a new perspective, we propose (b) *seeing from the hand's perspective*—camera mounted on the hand, which gets rendered as an electrotactile image on the back of the hand. (c) In our user study, we found that this enables flexible manual interactions, and supports ergonomic interactions, e.g., less crouching, leaning, craning, etc. (Photos with consent from participants)

Abstract. Sensory-substitution devices enable perceiving objects by translating one modality (e.g., vision) into another (e.g., tactile). While many explored the placement of the haptic-output (e.g., torso, forehead), the camera's location remains largely unexplored—typically seeing from the eyes' perspective. Instead, we propose that seeing & feeling information from the hands' perspective could enhance flexibility & expressivity of sensory-substitution devices to support manual interactions with physical objects. To this end, we engineered a back-of-the-hand electrotactile-display that renders tactile images from a wrist-mounted camera, allowing the user's hand to feel objects while reaching & hovering. We conducted a study with sighted/Blind-or-Low-Vision participants who used our eyes vs. hand tactile-perspectives to manipulate bottles and soldering-irons, etc. We found that while both tactile perspectives provided comparable performance, when offered the opportunity to choose, all participants found value in also using the hands' perspective. Moreover, we observed behaviors when "seeing with the hands" that suggest a more ergonomic object-manipulation. We believe these insights extend the landscape of sensory-substitution devices.

Keywords and Phrases: Haptics, Sensory substitution, Blind; Electrotactile, Wearable

1 INTRODUCTION

Perceiving the characteristics of objects (e.g., shape) at a distance is advantageous for preempting interactions (e.g., preparing grasp while reaching for an object), identifying parts of the environment (e.g., avoiding obstacles), and building spatial understanding. Neuroscientists have long established that during the reach phase of a hand grasping movement, humans (as well as other primates) "pre-shape the hand" [24, 64, 65] to best fit the object they intend to manipulate—these types of preemptive adjustments of one's grasp also led some to denote this phenomenon as anticipatory planning of reach-to-grasp movements [56]. Particularly, it has been understood that the target object's shape, size, and orientation influence the activity of hand muscles [16]. In fact, "vision appears to be more relevant for the final phases of the movement" [6], as one's hand approaches an object, real-time visual feedback becomes more critical to prepare their grasp accordingly [6, 18]. However, this is extremely difficult for Blind or Low-Vision individuals who cannot rely on sight for these adjustments during object manipulation.

Sensory substitution devices, while initially proposed to study brain plasticity, became powerful interfaces allowing users, especially those that cannot rely on vision, to *distally* perceive objects by translating information from one modality (e.g., visual) to another (e.g., tactile). Canonical examples of the many sensory substitutions in prior work include the *BrainPort* [7] and a

* Authors' pre-print of ACM CHI 2025 Paper

forehead device developed by Kajimoto, et al. [31], which stem from research in visual-to-tactile interfaces dating as early as the 1960s [3]. These interfaces, like others, utilize a camera for input and a tactile array for output. The camera is typically worn on the forehead and captures visual information from the *eyes' perspective*. Images are processed to extract features (e.g., contours of objects) and displayed to the user by means of a haptic device. Most commonly, the device renders the camera's view as a "tactile-image" using electrotactile [7, 27, 31] or vibrotactile [3, 42] feedback. Over the past decades, this type of sensory substitution has successfully enabled Blind, Low Vision, or blindfolded users to tactilely perceive many features of their surroundings from a distance [10, 51]. In fact, *BrainPort* has become a commercially available assistive technology. Given the success of sensory substitution, much research effort has been dedicated to design variations on these systems, especially focused on exploring which areas of the body to use for the haptic output. While the forehead [31, 52] and tongue [2, 27] are two of the most well-known candidates, other devices render their tactile-images to the user's abdomen [42], back [3], and even thigh [11].

Yet, while many have explored where to place the haptic-output, *the camera's location has remained largely unexplored*—with many sensory substitution devices using tactile-images from the *eyes' perspective*. Moreover, to match the viewing perspective, most devices utilize a haptic-output location with a similar frame of reference (e.g., similar viewing angle or even fully parallel) to that of the eyes, such as the case of the forehead (parallel to the eyes), tongue (same heading as eyes), and back/torso (mostly parallel, usually same heading). Intuitively, there are excellent design reasons to use this eyes-perspective and render tactile images to a body location with a similar frame of reference (e.g., forehead), namely the naturalness of the placement (i.e., head rotates, and the view rotates accordingly) as well as the view it affords (i.e., facing forward). These might explain why these devices are typically used for rendering surroundings (e.g., walking [42], avoiding objects [10], and so forth), but *rarely for assisting with interactions that involve object manipulation,* e.g., perceiving the affordance of the object (e.g., shape) in order to adjust hand shape for successful grasping [24, 64].

Hence, we explore **adding a new perspective for assisting with object manipulation**, distinct from the one afforded by the eyes, that might enhance flexibility of sensory substitution. To this end, we engineer & study a wearable device enabling users to **see with their hands** when hovering over objects. Figure 1 depicts users feeling tactile-images rendered on their hands from a camera mounted on the palmar side of their hands. This new perspective allows users of sensory substitution to leverage the hands' flexibility—hands move rapidly, reaching from multiple angles, exploring tight spaces, circling occluded objects, etc. To realize this, we implemented a novel sensory substitution device consisting of a wrist-worn camera, whose image is displayed as through a 5×6 electrotactile array on the *back* of the user's hand—moving the electrotactile array to the back of the hand prioritizes the ability to interact with physical objects with the palmar side of the hand.

To understand the benefits of "seeing with the hands," we conducted a study on Blind and Low Vision participants, as well as blindfolded sighted participants, who used the eyes' & hands' tactile-perspective, one at a time, to perform challenging manual tasks. We found that while both perspectives provided comparable performance, "seeing with the hands" resulted in more ergonomic interactions, especially when reaching for objects.

At this point, the reader might expect we are proposing replacement of the traditional (eye-view) sensory substitution with one that views from the hand's perspective. However, *this is not the case*. Our goal is to explore and understand the unique advantages afforded by "seeing with the hands" towards the goal of combining both approaches. In fact, we also had participants try out all interface combinations. We found that when given the option to use either or both devices, *all* participants chose to use *both*. We believe that this novel combination will unleash new modes of interaction and new benefits for users of future sensory substitution devices.

2 CONTRIBUTIONS & POSITIONALITY STATEMENT

Our contribution is the exploration of a novel interface concept for sensory substitution, in which users *see with their hands*, by feeling tactile patterns on the back of their hands captured from wrist-mounted cameras.

Benefits. Our approach has several key benefits: (1) it provides a fresh perspective to sensory substitution, by exploring a new location to place the visual & tactile components of the interface (camera on palmar side & tactile image on dorsal side); (2) the hands-perspective was found, in our study, to be suited for ergonomic reaching; (3) it enables new applications for sensory substitution, which we drew from participants feedback; (4) relocating the electrotactile array to the back of the hand provides a new design strategy for sensory substitution devices that wish to prioritize the user's dexterity; and, finally, (5) it is not a competitive approach to sensory substitution, we found that the hands-perspective can be easily combined with the traditional eyes-perspective, and, in fact, all our participants opted to do so in phase 2 of the study.

Positionality statement. Our device & study was co-designed and piloted by a blind lead co-author. This author was born legally blind and has no functional vision now (only light perception and color contrast). We acknowledge that this does not represent the lived experiences of congenitally blind (i.e., no visual memory) and low vision individuals. Moreover, as with our blind participants, our blind author has no prior experiences with sensory substitution, so design decisions were also made from a wish to improve the initial experience with these devices.

3 RELATED WORK

The work presented in this paper builds on the field of haptic devices for sensory substitution. Since our goal is to support users wishing to non-visually explore and interact with anything in their surroundings by means of tactile sensations, we primarily focus our related work on tactile-visual *sensory substitution devices*. We also succinctly overview devices for haptic guidance, especially, those also exploring a hand-perspective. Finally, given that our implementation is based on electrotactile, we succinctly review this haptic technique.

3.1 Sensory substitution

In 1969, Paul Bach-Y-Rita developed the first sensory substitution device, *Tactile Television*, which converted images captured by a stationary camera into tactile feedback on the person's back [3]. After extensive training, blind individuals were able to understand the movements of people and objects in the environment, etc. Since this pioneering work, many sensory substitution devices have been developed. While the original device enabled a visual-to-tactile translation, others have explored translating to other senses (e.g., visual-to-auditory substitution [12, 44, 60]). Given the extensive range of this field, readers can refer to the reviews on the subject [5, 15, 36, 40, 62].

When focusing on visual-to-tactile substitution, tactile images have been rendered to the forehead [31, 52], tongue [2, 10, 27], abdomen [32, 42], back [3, 23], and thigh [11]. These devices often capture visual information from the *eyes' perspective* or similar references (e.g., torso) and render it to a tactile array using vibrotactile [3, 42] or electrotactile [2, 31]. A modern example is the *BrainPort* [7], a commercialized product featuring an electrotactile display on the tongue. In most cases, the image is processed to extract features—typically, contours—that are rendered as tactile sensations. For instance, if a person using *BrainPort* or the forehead device proposed by Kajimoto, et al. [31] "looks" at a door, they will feel a rectangle of tactile bumps on their tongue or forehead.

Others explored capturing information from the perspective of body parts other than the eyes or the torso. In audio-visual substitution, Brown, et al. [9] found it was easier to recognize objects via a handheld camera (like a flashlight), compared to using a head-mounted camera. In tactile-visual substitution, *FingerSight* [21] proposed a finger-mounted camera that captured edges to be perceived on the finger via two vibromotors. Krishna, et al. [35] used 14 vibromotors on the back of fingers to present facial expressions. *ThroughHand* [26] engineered a tabletop device for visually impaired users comprised of an overhead camera and a shape-changing display; by resting their palms on the surface, users are able to feel the content (e.g., 2D video games) as the pins of the shape display update—while designed for a purpose very different from our approach, this interface shares one common goal with ours, i.e., rendering multiple stimulation points on the user's hands. Kilian, et al. [33] translated the depth image of a camera mounted on the back of the hand to a tactile pattern on a 3×3 vibrotactile array, enabling blind participants to navigate an obstacle course. Lobo, et al. [41] used a line of vibromotors on the legs to represent the height of upcoming obstacles. *SpiderSense* [43] explored tactile perspectives from multiple parts of the body, by translating distal information to servo motors that push against the user's skin.

Hands have been shown to be effective locations for perceiving tactile images. Yet the aforementioned sensory-substitution systems mostly focus on *perceiving virtual images* (e.g., *ThroughHand* [26] renders a game screen, wearable gloves [35] render emoji icons) or *navigating the environments* (e.g., *Unfolding Space Glove* [33] assists only with avoiding obstacles, of identical shapes, while walking). As such, existing sensory-substitution systems rarely consider *interactions with physical objects* (e.g., prepare grasp for object's affordance [24, 64]). In contrast, hand-worn haptic interfaces have been explored extensively to guide the users' hand to interact with objects, which we discuss next.

3.2 Haptic guidance from the hands' perspective

Researchers have explored haptic cues to guide the user's hand closer to a target object. Such haptic patterns are typically designed to be perceived from the *hands' perspective*—the spatial information of the target is *relative* to the hand. This has been shown to be an intuitive strategy, e.g., if the target is on the left to the user's hand, the left vibromotor on the hand [8, 19] or on the wrist [48, 63] will vibrate to guide the user to move to the left.

While many works were realized in virtual environments, a follow-up work of *FingerSight* [49] contains a miniature camera with four vibromotors worn around the index finger to indicate the direction to a target. *PalmSight* [68] used a depth camera placed on the palm and five vibrotactile motors on the back of the hand. The direction and distance of a target object (from the depth camera) relative to the hand were translated to activate corresponding vibromotors. While the authors described their work as sensory substitution, they emphasized that *PalmSight* "relies on the computer to make high-level judgement, e.g. whether the target object is identified and what its relative location is to the hand" [68]. This highlights the core difference between *haptic-guidance* and typical *sensory-substitution* systems—haptic-guidance systems must be able to *track the object of interest* which relies on the assumption that (1) the user has indicated an object (they assume it exists in the scene); and, (2) the system will track this object for the user. With these assumptions in place, the system then resorts to different haptic cues to steer the user closer to the tracked object. Compared to haptic guidance, tactile sensory substitution foregoes these assumptions and lets users non-visually parse the scene by themselves—users do not indicate objects of interest or ask the system to track objects. Instead, they receive information about their surroundings and make judgements by themselves (decisions happen in the user's brain, not in the computer).

3.3 Bringing the hands' perspective to sensory substitution for manual interaction

Instead of guiding to objects, De Paz, et al. [13] explored a sensory-substitution device that allows free exploration from the hands' perspective to assist non-visual grasping. The device consists of two vibromotors worn on the index finger and thumb. The intensity of the vibromotors increase as fingers approach objects (akin to a game of "hot cold"). The study shows that blindfolded participants were able to locate, identify, and grasp cylinders on a table in a fully-tracked environment using motion-capture system. Yet, since their device only featured two haptic stimulation points (two motors), the authors reported that the device fell short on presenting the shapes of objects [13].

We see a missed opportunity here— how can we leverage the *hands' perspective* to support the complete interactions involved in object manipulation (e.g., including shape recognition)? We believe that by bringing more expressive sensory substitution to the hand (i.e., 2D tactile display allowing to feel tactile images), we can uncover the unique benefits offered by the hands' flexibility and mobility to assist object manipulation.

3.4 Electrotactile stimulation

Electrotactile stimulation is a technique that creates tactile sensations by means of electrical impulses, delivered across electrodes at user's skin [30, 55]. Electrotactile has been shown to generate various sensations on the skin (touch, pressure, textures) [17, 58], and offers several advantages over canonical vibrotactile feedback. First, since electrodes can be made thinner (just 0.1 mm thick) than mechanical actuators (physical displacement requires space), electrotactile arrays can be made slimmer and more conformable than a vibrotactile arrays and therefore suitable to be worn on various parts of the body. Second, electrotactile feedback has been shown to be felt more localized than vibrotactile feedback [53, 61], which makes electrotactile a suitable method for high-resolution tactile arrays. As such, besides sensory substitution [28], there is growing interest in electrotactile for many interfaces, such as—touch feedback in virtual environments [57, 59, 66, 67], guidance displays on the user's wrist [53] and foot [61], and prosthetics [54]. We invite the reader to refer to [34] for a thorough review of electrotactile and its applications.

4 A NEW PERSPECTIVE FOR SENSORY SUBSTITUTION TO ASSIST WITH MANUAL INTERACTIONS

Typical sensory substitution interfaces are used for assisting with perceiving one's surroundings (e.g., navigation, avoiding obstacles, etc.), which has led to the camera's most common position at the eyes (and in some work, also at torso, waist-level). While prior work explored placing camera on the back of the hand as to avoid obstacles [33], this leaves us to wonder: *could a more flexible perspective, i.e., facing the direction of a possible hand grasp, be useful?*

4.1 "Seeing with the hands" for manual interactions with physical objects

We explore a new tactile-perspective by which users of sensory substitution devices "see with their hands"— feel tactile-images rendered onto their hands, which are captured from hand-mounted cameras *on the palmar side*—This is the side of the hand facing towards objects to grasp for hand manipulation (as opposed to [33]). Figure 2 illustrates our concept by contrasting it with the more traditional eyes' perspective: (a) rather than having a camera seeing from the eyes' perspective and a tactile interface to feel via the eyes' frame of reference, we explore (b) *seeing with the hands* via a tactile interface attached to the back-side of the hands—this allows users to preserve tactile sensitivity on the palmar side to grab and manipulate objects with dexterity. This perspective is unique in that it enables users of sensory substitution to leverage the hands' affordances—namely their flexibility & speed as hands can move rapidly around the body, skirting objects, reaching from multiple angles, getting into tight spaces, circling behind occluded objects, etc. As depicted in this example (from our user study), users can use the "hands view" to perceive the shape of the object and adjust their grasp during reaching, even when manipulating a risky object (e.g., a soldering iron).

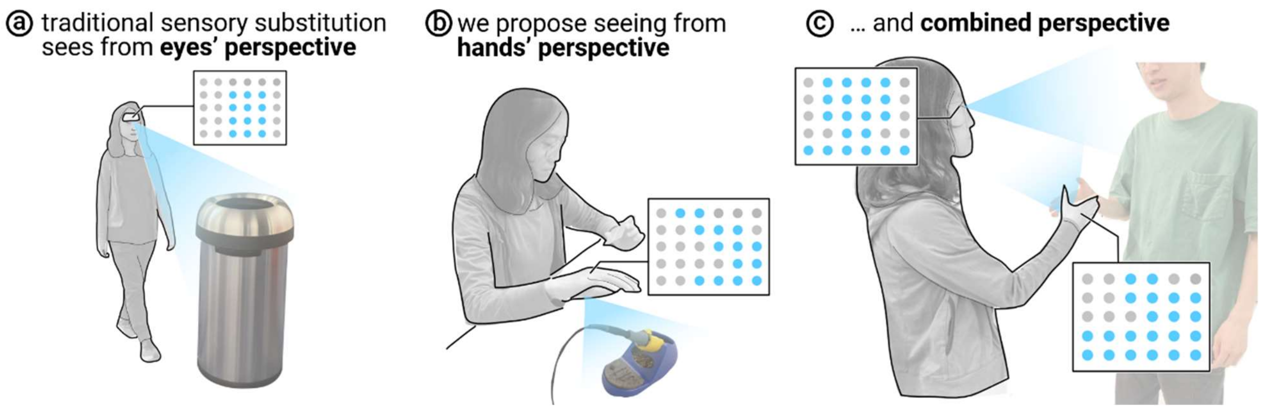

Figure 2: Contrasting three different tactile-perspectives for sensory substitution: (a) eyes; (b) hands; and (c) combined.

Besides contrasting our approach with the traditional (eyes) perspective, Figure 2 (c) highlights an important aspect of our concept: we are *not* proposing to replace the eyes' perspective with that of the hands'; instead, we believe the advantages afforded by each way of seeing allows these approaches to *combine*. i.e., by seeing from eyes, hands, or both. In fact, we found in our study that when given the option to use either or both devices freely, *all* participants used *both*.

4.2 Implementation

To instantiate our concept, we implemented a wearable prototype. To help readers replicate our prototype, we provide the necessary technical details. Additionally, all source code & materials will be made publicly available[1].

Figure 3 (a) depicts our prototype worn at the hands' perspective. For the purpose of our study, we also adapted this prototype to the eyes' perspective, by moving the camera to the forehead (on a glasses' frame) and the electrotactile to the forehead (mounted on a headband as in [31, 52]). Regardless of the type of perspective, our prototype is comprised of two main modules (vision & tactile) connected to a PC where the processing is performed.

[1] https://lab.plopes.org/#seeing-with-the-hands

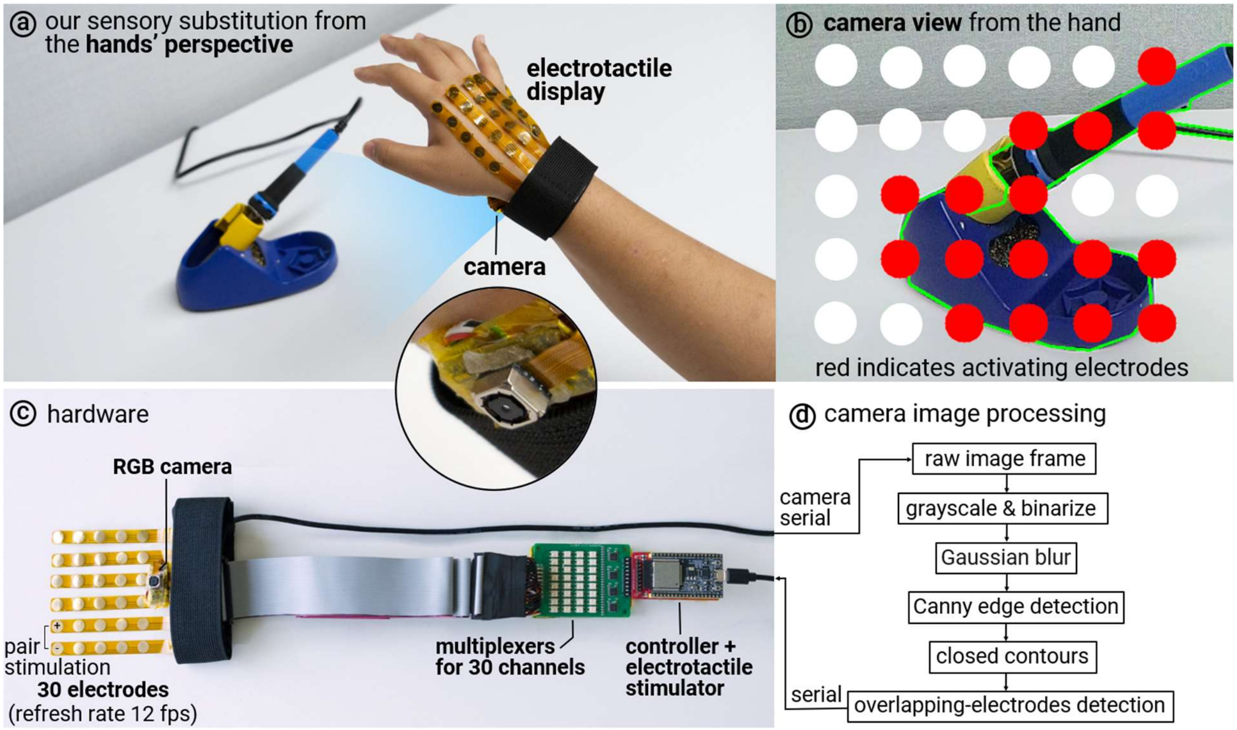

Figure 3: Implementation: (a) overview of "seeing with the hands"; (b) camera view after image processing, where the circles represent the corresponding electrodes on the back of hand; (c) hardware components; and, (d) image processing pipeline.

Vision module. We utilize a miniature camera (10×10×5 mm) with 60° field of view, to minimize obstruction, especially when mounted on the hand. The camera sends its data over USB (15 FPS). A Python program uses OpenCV to process images, using the simple pipeline in Figure 3 (d). First, we threshold the raw RGB image to grayscale and binarize (at a threshold of 90, adjustable for lighting conditions, albeit not automatically in our implementation). To reduce noise, we apply a Gaussian blur (5×5 kernel). Then, contours are detected with the Canny edge detector [46] and filtered based on their area, retaining only those greater than 1,000 pixels. A final polygonal approximation [47] is used for contour refinement. Finally, a grid of circles (5×6) is projected on top of the processed image, each circular-cell depicting an electrode on the user's skin. An electrode on this grid is considered activated if a contour passes inside, as depicted in Figure 3 (b). The list of activated electrodes is transmitted via serial communication to a microcontroller.

Tactile module. Our implementation makes use of electrotactile stimulation. The hardware is depicted in Figure 3 (c). An electrotactile stimulator [29] and our multiplexer are controlled with an ESP32 microcontroller board. Our multiplexer (similar architecture as [61]) routes which electrodes outputs the stimulator's signals to the user's skin. It can route one signal to a maximum of 32 electrodes. The tactile arrays were fabricated using flexible PCBs (flexPCB), since their polyimide substrate is strong (e.g., hard to rip), while still being relatively thin (0.1 mm). 30 electrodes (⌀8mm) are used to cover the back of hand or the forehead placed in a 5×6 grid with equal spacing (15mm), which is larger than the two-point discrimination on the back of hand (9mm [50]) and forehead (3mm [31]).

Stimulation parameters. We use a pulse generator with programmable current output (circuit design from [29]). For each tactile pixel, we program the circuit to form an electrode pair (we found in pilot experiments that the sensation was robustly felt at the ground electrode, despite the location of the positive electrode; thus we chose to stimulate a horizontal pair of electrodes as depicted in Figure 3c). We stimulate with a square-waveform with a pulse width of 360 µs at 200 Hz. These values were determined through pilot tests as they produced localized and comfortable sensations on the back of the hand. The current (1-5mA) is calibrated for each user (see *User Study* for calibration details). We utilize time division for stimulating multiple tactile pixels. The refresh rate of the entire tactile array is 12 frames per second.

5 USER STUDY: UNDERSTANDING THE CONTRIBUTION OF THE HANDS' PERSPECTIVE

The goal of our study is to understand whether there is a unique contribution of seeing with the hands for tactile-visual sensory substitution. Therefore, we designed a study with two phases: **(1) single-perspective phase**: where participants completed tasks using either the hands- or eyes-perspective, but not both simultaneously. This was purposefully designed to collect data (quantitative, qualitative, and observational/behavioral) that captured where they succeeded or struggled with the affordances of each device **(2) combined-perspective phase**: where participants completed a final task in which they could freely choose which perspective they use (eyes', hands' or both at the same time).

Since our goal is to gain insights that might one day impact users of future substitution devices, most of whom are Blind or Low Vision, our study was co-designed and piloted iteratively by one of our blind lead authors.

This study was approved by our institutional ethics committee (IRB21-1229).

5.1 Tactile perspectives (sensory substitution interfaces for our study)

Hand's perspective (*hand-device*): This is our proposed new perspective. This was implemented by means of the device described in *Implementation.* Participants wore the *hand-device* on their dominant hand alongside its back-of-hand electrotactile display that renders tactile image from the wrist mounted camera.

Eyes' perspective (*eye-device*): This is a baseline condition that we chose to represent the traditional approach, with the camera mounted at eyes' level (between the eyes on the frame of an empty glasses). We chose the forehead from prior work (e.g., [31]), as the forehead was shown also be suitable for electrotactile display.

Apparatus. Besides the location of the camera/tactile-array, both devices were identical in their implementation (same hardware & algorithm). Participants also wore both (eye- & hand-) devices at all times. The study was conducted in a room with white walls. A table was used to place objects. For data collection, a fisheye camera was mounted in front of the table. HTC VIVE Trackers were attached to the participants' dominant hand and head for tracking trajectories.

Minimizing bias. Importantly, all participants had no prior knowledge about sensory substitution devices and were not told which was our interface condition (*hand-device*) and which was the traditional sensory substitution device (*eyes-device*), instead they were neutrally asked to try both.

5.2 Participants

Eight participants were recruited, five were male and three were female (average age=36 years, SD=15.23). Four were sighted (PS1-4) while four were Blind or Low-Vision (PB5-8). Participants were offered the option not to have their videos recorded, and two participants preferred to not be recorded. Participants were compensated with 50 USD.

6 CALIBRATION OF ELECTROTACTILE INTERFACES & TUTORIAL

Before the trials, we calibrated both electrotactile displays and provided an explanation on sensory substitution.

Calibration. An iterative calibration of all 60 electrodes (forehead & hand arrays) was performed to ensure that each of the electrotactile sensations generated by the array could be felt clearly and localized. During calibration: (1) each tactile pixel (an electrode pair) was stimulated; (2) participants then verbally assisted the experimenter with adjusting the intensity of the stimulation (starting from 0mA and increasing by 0.5mA steps), until; (3) the stimulation at the target location was felt clearly and without causing pain; (4) finally, if the sensation was not collocated with the electrode pair (e.g., causing referred sensation at the fingers), the electrode pair was skipped to avoid confusion (at most we only allowed to skip five pairs out of 30 per participant, to ensure at least 25 active and well-calibrated electrodes)—this calibration process is typical in electrical stimulation devices (e.g., similar to [25, 57]).

Tutorial. Most studies on sensory substitution use long training phases, sometimes up to several hours [5]; however, we wanted to explore how participants might make use of natural affordances of each interface so we limited this to 10 minutes per condition (order counter-balanced). In these tutorials, participants had a chance to try sensory substitution for the first time (even our Blind and Low-Vision participants had never experienced such devices) and also experience how electrotactile feels. Participants were asked to: (1) use the device to feel a sponge ball without touching it—this allowed participants to get familiar with field of view of the camera; (2) trace the outline of a plastic frame—get familiar with feeling a bigger object containing line and corner features; (3) explore a PET bottle—get familiar with objects that have a significant third (height) dimension; (4) find & grasp the sponge ball three times—this allowed them to get familiar with the mechanics of the upcoming trials.

6.1 Study Phases

Phase 1: single-perspective phase (comparison of eyes' vs. hands'). After training, participants were asked to use each perspective (eyes or hands condition) *one at a time* to complete four tasks while blindfolded (regardless of visual acuity). The tasks were designed by our lead blind author so as to involve a diverse set of everyday manual interactions (extending prior visual-tactile sensory substitution which studied on simple objects [13] or navigation [33]). The tasks as depicted in Figure 4, are: **(1) object identification task**: finding and picking up a pen among three objects on the table (similar to [1] studying daily objects); **(2) object orientation task**: picking up a "hot" soldering iron by its handle (the iron was not actually hot to ensure their safety), which represents a safety task with similar concepts in [38]; **(3) hand-eye coordination task**: find and pick up a bottle lid from the table, and subsequently find a bottle on the table (without touching the bottle, align the two, and aim the lid at the bottle's opening, screw the lid on the opening (this task is similar to a lid-aiming task in [37]); and, **(4) obstacles & occlusion task**: find a small notebook inside one of the three boxes that are placed unknowingly beforehand, without picking up the wrong object (a pen is placed inside another box). The final task represents a scenario with obstacles which were shown to affect reaching [45].

Figure 4: (a) Phase 1 only allowed participants to use one device at a time to complete the following tasks: (1) object identification; (2) object orientation; (3) hand-eye coordination; (4) obstacles & occlusion. (b) Phase 2 allowed them to try any combination of devices (hands, eyes, or both simultaneously) to find a person and shake their hand.

Trial design. We limited each trial (i.e., a task done using one condition) to a maximum of five minutes—if by the end of the time participants were not able to complete it, we moved to the next. To ensure that participants solved the tasks *using the actual sensory substitution interfaces* and not just by touch alone, we instructed them that if they touch an incorrect object, this counted as a mistake. This encouraged that they explored objects via the sensory substitution prior to attempting to grab them. For instance, if a participant in task 1 (picking up a pen among other objects) grabbed the wrong object, a mistake was counted, and the objects were reshuffled to new locations (the completion time was paused while experimenters reshuffled). The participants were given time for breaks in between tasks.

Questionnaire & metrics. During each trial (a task, performed once per device), we collected videos of their movements, completion time, hand and head trajectories, and number of mistakes. Finally, once they completed a trial, they were asked to rate physical and cognitive load (these two items were taken from the NASA TLX [20]) as well as provide comments on what they experienced, which were transcribed by an experimenter. Observed behaviors were transcribed from videos; two of the authors annotated each recorded video using a sequences of codes [39] which included (1) descriptions of hand & head movements and locomotion, and (2) the occurrences of un-ergonomic postures, e.g., neck inclination, trunk inclination, and crouching according to ISO 11226 and EN 1005-4 [14].

Condition order. Within each task, the condition order was counter-balanced across participants (i.e., if a participant used the hands-perspective first for task 1, they would use the eyes-perspective first for task 2).

Phase 2: combined-perspective phase: Finally, participants completed a task in which they could choose any of three perspectives: **(1) eyes-only, (2) hand-only, or (3) both** at the same time. To toggle between the three different perspectives, they simply said the desired perspective out loud ("hand", "eyes", "both") and an experimenter switched them at a press of a button. **The task was to find a person and shake their hand** (extending a task similar to [42] with additional manual interaction). They were told that the person could be standing at any location of the room with their hand extended, including higher or lower than a normal handshake position. While at first glance this task seems easier than our previous ones (e.g., than finding an object in boxes), pilots with our Blind author confirmed this task is challenging. First, this task is less spatially constrained, i.e., while the objects were on an unmovable table (acts as a frame of reference), a person can stand anywhere in a room (larger frame of reference). Secondly, moving and exploring the room in search of a person is harder than finding objects on an empty table (clear signals), since a person can stand behind camera tripods, in corners, etc. This heightened difficulty was intentional since we wanted to see what participants used each perspective for. At the end of this task, experimenters asked the participants to explain their rationale when choosing the perspective(s) they used.

6.2 Results from interactions using a *single* tactile perspective at a time (eye or hand phase)

We first report findings from the first phase in which participants used one device at a time for each task. The quantitative results including cognitive & physical loads, mistakes, and task durations are reported in Table 1.

Comparable performances and loads for *hand*- and *eye-device*. We observed a comparable average number of mistakes for both devices, with the *hand-device* at 0.7 (SD=1.0) and the *eye-device* at 0.7 (SD=1.0). Across both perspectives, five (out of a total of 32) tasks were not completed within the limited time. The task durations for the hand-device is in average 148 seconds (SD=102), and the eye device was 139 seconds (SD=97).

The average physical load was lower with *hand-device*. While we did not find a statistical difference in cognitive load (hand: AVG=4.2, SD=1.7; eye=4.4, SD=1.7), we found that the average physical load with hand-device was significantly lower than of the eye-device (AVG=2.1, SD=1.2; eye: AVG=3.0, SD=1.9; paired t-test; p<.001, F(31)= 3.69).

		cognitive load (7-point)			physical load (7-point)			mistakes			task duration (sec)		
	tactile device	sighted	BLV	all	sighted	BLV	all	sighted	BLV	all	sighted	BLV	all
task 1	eye	5.8 (1.0)	5.0 (1.4)	**5.4 (1.2)**	4.3 (2.1)	2.0 (1.4)	**3.1 (2.0)**	0.8 (0.1)	1.5 (1.3)	**1.1 (1.1)**	233 (114)	118 (75)	**175 (109)**
	hand	4.0 (0.8)	4.5 (1.7)	**4.3 (1.3)**	2.5 (1.3)	1.8 (1.0)	**2.1 (1.1)**	1.0 (0.8)	1.8 (1.5)	**1.4 (1.2)**	160 (103)	179 (140)	**170 (114)**
task 2	eye	4.8 (0.5)	3.5 (2.1)	**4.1 (1.6)**	3.5 (1.7)	2.3 (1.9)	**1.9 (1.8)**	0.3 (0.5)	0.3 (0.5)	**0.3 (0.5)**	114 (44)	57 (52)	**86 (54)**
	hand	4.5 (1.7)	4.3 (2.8)	**4.4 (2.1)**	2.5 (1.0)	2.0 (2.0)	**2.3 (1.5)**	0.3 (0.5)	0.5 (1.0)	**0.4 (0.7)**	165 (101)	100 (134)	**132 (115)**
task 3	eye	3.8 (1.0)	2.8 (2.2)	**3.3 (1.7)**	4.3 (2.2)	1.8 (1.0)	**3.0 (2.1)**	0.8 (1.5)	0.8 (1.0)	**0.8 (1.2)**	158 (101)	99 (114)	**128 (105)**
	hand	3.8 (1.0)	4.0 (2.4)	**3.9 (1.7)**	2.8 (1.7)	1.3 (0.5)	**2.0 (1.4)**	0.8 (1.5)	0.8 (0.5)	**0.8 (1.0)**	152 (101)	114 (119)	**133 (104)**
task 4	eye	5.0 (0.8)	4.5 (2.6)	**4.8 (1.8)**	4.3 (2.1)	1.8 (1.0)	**3.0 (2.0)**	0.3 (0.5)	1.0 (1.4)	**0.6 (1.1)**	178 (103)	158 (115)	**168 (102)**
	hand	4.5 (1.7)	4.0 (2.2)	**4.3 (1.8)**	2.8 (1.0)	1.5 (0.6)	**2.1 (1.0)**	0.5 (1.0)	0.3 (0.5)	**0.4 (0.7)**	138 (73)	172 (108)	**155 (87)**
all	eye	4.8 (1.0)	3.9 (2.1)	**4.4 (1.7)**	4.1 (1.8)	1.9 (1.2)	**3.0 (1.9)**	0.5 (0.9)	0.9 (1.1)	**0.7 (1.0)**	170.6 (95.5)	108 (91)	**139 (97)**
	hand	4.2 (1.3)	4.2 (2.1)	**4.2 (1.7)**	2.6 (1.1)	1.6 (1.1)	**2.1 (1.2)**	0.6 (1.0)	0.8 (1.0)	**0.7 (1.0)**	153.8 (85.6)	141 (118)	**148 (102)**

Blind & Low Vision (BLV) participants rated **lower physical loads** for both devices

overall, ***hand-device*** was rated **lower physical load** than *eye-device*, with **comparable performance**

Table 1: Phase 1 study results (Numbers are average with SD in parentheses).

Blind & Low Vision (BLV) rated lower physical load. We found a significant difference between BLV and sighted participants in physical load with eye-device (BLV: AVG=1.9, SD=1.2; sighted: AVG=4.1, SD=1.8; paired t-test; p<.001, F(15)=5.03), and with hand-device (BLV: AVG=1.6, SD=1.1; sighted: AVG=2.6, SD=1.1, paired t-test; p<.05, F(15)= 2.47). We did not find a significant difference regarding cognitive load. Moreover, we observed comparable mistakes and duration.

Emergent scanning behavior for exploring objects. We observed that participants adopted scanning movements (despite never being told about this)—move their eyes-device or *hand-device* back and forth to "scan" objects. Also, they often reoriented the devices to scan objects from different angles —by rotating their hand when using the *hand-device*, or rotating the whole head/torso/body when using the *eye-device*. This scanning behavior was confirmed by our trajectory data. We found that the average trajectory length showed more movement of the body part where the device was placed. When using the *hand-device*, the hand moved an average of 11.8 m (SD=3.4) while the head moved 5.6 m (SD=1.3). When using the *eye-device*, the hand moved an average of 7.7 m (SD=2.1) and the head moved 9.1 m (SD=2.4). As we will see next in the video observations of participants' behaviors, the length of the head movements (almost comparable to those of the hands) were felt as less ergonomic than hand movements.

***Eye-device* led to more unergonomic behaviors than *hand-device*.** When using the *eye-device*, crouching was more commonly observed to perceive the object from a different angle or to avoid occlusion (e.g., see inside of boxes). Across all 32 trials, 17 crouches were observed with the *eye-device*, compared to only six when using the *hand-device* (an example shown in Figure 5b in thumbnails #3 and #4, which is considered an awkward posture [14]). This observation was corroborated in participants' recounts of their experience. For instance, PS2 stated "[with *eye-device*] I need to bend down, and it was hurting a bit". Similarly, PS3 stated "[with *eye-device*] you can't scan it the same way as using your eyes (…) having to crouch and squat to find these objects". Other unergonomic behaviors [14] while using the *eye-device* were observed for all participants, such as craning the neck (i.e., neck flexion). Furthermore, across all trials, we observed 18 trunk forward inclinations (an example shown in Figure 5b in thumbnail #2) while leaning when using the *eye-device*, compared to only six with the *hand-device*. To this end, PB6 stated: "I don't think [eyes-device] is practical because it's a pain in the butt to always crane your neck to figure out what it is". In contrast, some participants commented on the *hand-device* to feel freer. Namely, PS4 stated "hand[-device] is freer and easier to search". Similarly, PS3 stated "I feel like I really prefer the hand[-device] for scanning and the head[-device] for trying to tell like the shape of the object". Exemplar behaviors in the study are depicted in Figure 5.

Figure 5: (a) When using the *hand-device*, participants explored the objects by scanning with their hand. (b) The *eye-device* led to unergonomic behaviors such as crouching and craning necks. (Photos with consent from participants)

***Hand-device* felt easier.** Namely, across all their feedback, we found 15 trials where, without being prompted, participants specifically stated that the *hand-device* was easier to use, and only 2 trials where they specifically stated the *eye-device* was easier (for the remainder trials, no specific device was stated to be easier). While there was a high degree of inter-task agreement (i.e., PS1, PS2, PS4, PB6, and PB7 always specifically stated that *hand-device* felt easier regardless of the task they commented on), there was one preference that was task dependent (i.e., PS3 specifically stated they preferred the *eye-device* for task 2, but the *hand-device* for task 1).

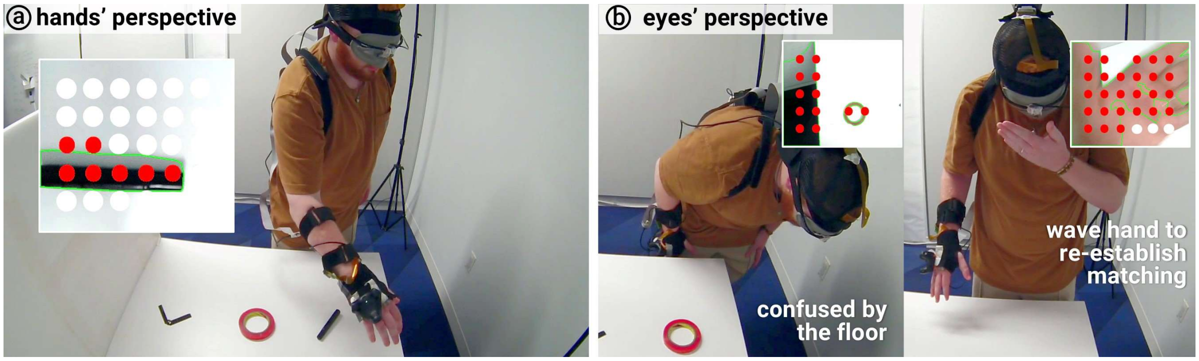

Figure 6: Contrasting the frame of reference of (a) hand moving over the table, and (b) head scanning over the table, which caused this participant to get lost and try to wave their hand to re-establish a reference. (Photos with consent from participant)

***Eye-device* requires spatial reference.** We observed cases in which participants momentarily were lost with their exploration, i.e., not knowing where they were with their device. While this happened for both devices, participants exhibited different behaviors for each device. Figure 6 (b) depicts one example of this, in which PS3 is momentarily lost when using the *eye-device*. They searched for the edge of the table, but confused by their spatial understanding, they waved their hand in front of the eye-camera to re-establish the matching between the tactile image and the physical world. PS3 stated: "I kept finding the floor because of the color contrast". Three participants directly commented on this difficulty in finding their frame of reference. To this end, PS1 stated: "It was easier to figure out where the [hand] camera is intuitively, all I had to do was imagine a camera on the wrist, like, when I think about touching things it needs to go to my palm like having eyes on my wrist it made it a lot easier." Later, they contrasted this with their experience with the *eye-device*, stating: "I map out range that I can see, I used [the] side of the table and feeling of continuous [stimulation] as a reference for where table started [and] ended" (by physically moving their head along the edges of the table). Finally, PS2 stated: "I actually had really hard time [locating] where it is, so I tried to [zoom to] table and maybe it was just near the edge. It was very confusing. I felt like maybe it would be an edge but maybe it would be object also".

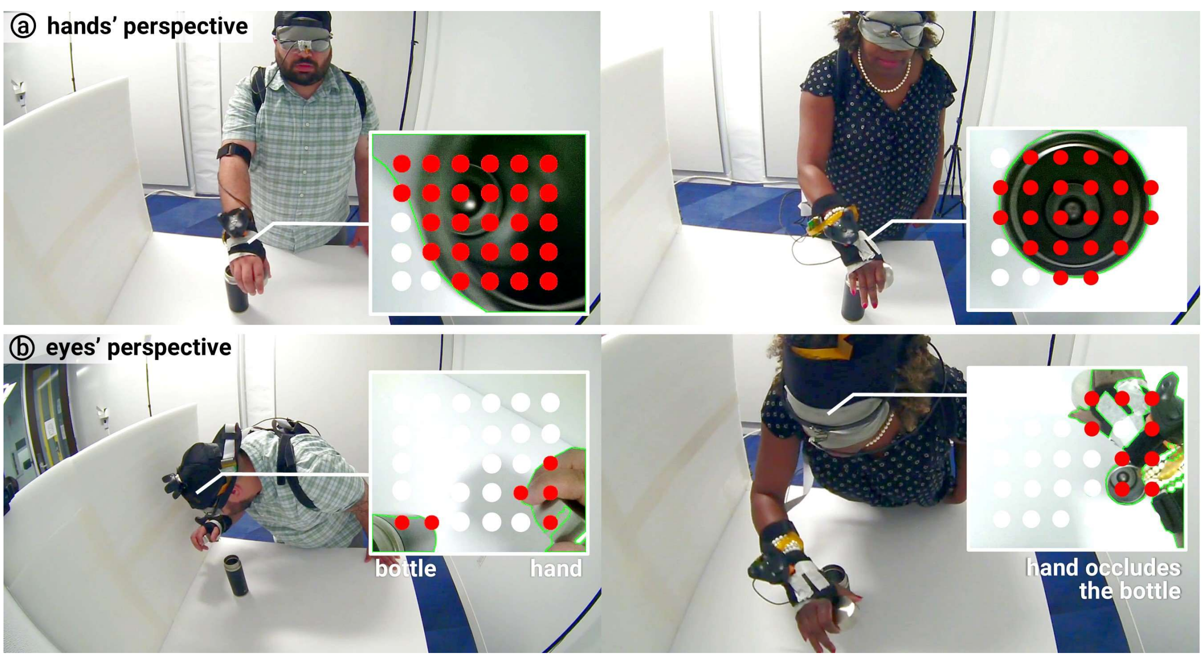

Figure 7: Participants used the (a) *hand-device* to adjust their reaching trajectory by keeping the object at center of the tactile array; and used the (b) *eye-device* to align the lid and the bottle. (Photos with consent from participants)

Reaching for objects. Generally, we observed that the *hand-device* allowed for a smoother pursuit when reaching objects, while reaching for objects with the *eye-device* required additional matching and was subject to occasional occlusions from the hand during reaching. Figure 7 illustrates some of these observed behaviors: (a) when using the *hand-device*, we observed how participants performed their reaching gestures, often by keeping the object in the center of the tactile array and then pursuing it; in contrast, (b) when using the *eye-device*, participants fixed their head orientation and then moved their hand into the view, checking once it overlapped with the object which indicated the hand was aligned with the object. While this behavior often ran into hand occlusions and can cause confusion (e.g., PB3 overshot their hand to target), it can also be beneficial, for instance, for alignment tasks, such as putting the lid on the bottle as shown; in fact, PB5 stated they found this easier for the bottle task and we observed PS4 using a similar strategy for the soldering iron. From their behaviors we also observed that both PS4 and PB5 were at times confused by their own hand occluding the view, despite the fact they were able to create a compensatory strategy and make it work. Specifically, regarding reaching and alignment with objects, PS1 stated: "[with eye-device] I could feel where the bottle cap was and match where my hand was on the head, [it] helped me pinpoint where I should be moving". Similarly, PS4 stated: "I changed strategy for putting on the lid [after a mistake]. I placed my head so that the bottle is on the left side of the forehead. I tried to approach it with my hand, and it worked".

6.3 Results from allowing participants to *choose* any tactile perspective (eye, hand, both—phase 2)

Finally, we report the results for the second phase of the study, where participants were allowed to choose any combination of tactile perspectives (eye, hand, or both at the same time) to complete the handshaking task. The quantitative results are reported in Figure 8, particularly a per-participant timeline of device usage in this final phase.

Task difficulty. One participant was not able to find the extending hand within the allocated five minutes. The average cognitive load was 4.3 (SD=2.6) and physical load was 2.9 (SD=2.2), both slightly higher than in the average of all the individual tasks from phase 1, which was expected since this task was less constrained than that of phase 1.

All participants used both devices. For this task, participants were able to freely choose which device to use (task started with none selected). Overall, we found that all participants used both devices at least once to solve the task. Specifically, two participants chose to use both devices at all times (PB5, PB8). Three participants switched back-and-forth between the *eye-* and *hand-device,* but only one at a time (PS1, PS2, PB6). Finally, the three remainder participants chose to use all devices either individually, or at the same time (PS3, PS4, PB7). The aggregated timelines of device use are shown in Figure 8.

cognitive load (7-point)			physical load (7-point)			mistakes			task duration (sec)		
sighted	BLV	**all**	sighted	BLV	**all**	sighted	BLV	**all**	sighted	BLV	**all**
5.5 (1.7)	3.8 (1.8)	**4.6 (2.6)**	4.0 (2.3)	1.8 (1.0)	**2.9 (2.0)**	0.3 (0.5)	0.3 (0.5)	**0.3 (0.5)**	208 (93)	91 (43)	**150 (92)**

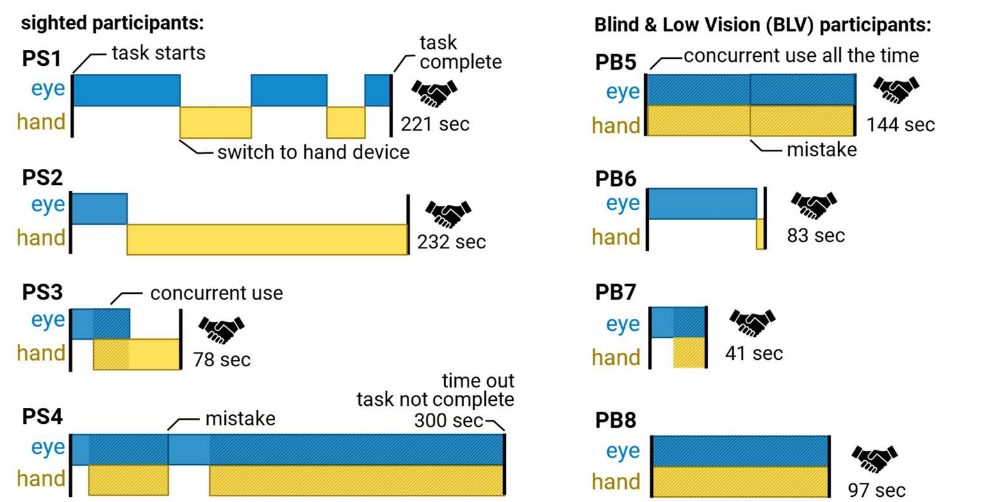

Figure 8: Results from phase 2: (top) Summary of cognitive load, physical load, mistakes, and task duration (Numbers are average with SD in parentheses); (bottom) timeline of device use for the final task (handshaking).

Comparing Blind & Low Vision (BLV) with sighted participants. We found that BLV participants took, on average, less time to finish the task when compared to sighted participants. In terms of cognitive and physical loads, we did not find a significant difference. Interestingly, the two participants who chose to use both devices at the same time were BLV participants (PB5, PB8). Specifically, two sighted participants (PS1, PS2) mentioned it felt overwhelming to use both devices concurrently; conversely, this type of comment was absent with BLV participants.

Switching & concurrent perspectives. As aforementioned, the majority explored all combinations to solve this task, including both trying the devices individually or concurrently. Some participants stated the rationale behind the strategy naturally occurred to them while using the devices in combination. To this end, PS4 stated they used a switching strategy as it helped with attention: "I was not using both, I was focused to either one (...) it was really switching, I used my head[-device] to detect a big object and hand[-device] to detect the arm [of the person to handshake] (...) switching the focus to one [device] was not too hard." To this, but using the devices concurrently rather than switching, PB8 stated: "I was using both equally. I like confirmation from both of them".

***Eye-device* to find the person vs. *hand-device* to find extended hand.** We found that for participants who used a single device at the start of the task (PS1, PS2, PS3, PS4, PB6, PB7) chose to use *eye-device first*, and switch to the *hand-device later*, suggesting a strategic use of devices (as shown in Figure 8). We observed, as exemplified in Figure 9, that participants naturally used the devices for different purposes. The *eye-device* was employed for finding the person in the room (in the words of PS4, "[the] big thing"). Conversely, once they located the person, they focused/switched to the *hand-device* to find the person's extended hand (in the words of PS4, "[the] detailed part"). All but one participant used this strategy, which suggests they naturally found the *eye-device* to provide a sort of overview, while the *hand-device* to provide a sort of pan & zoom—their feedback further corroborated this. For instance, PB6 stated: "I was focused on the [eyes-device] giving me direction on where he is and I was focused on hand[-device] to find his extended hand (...) once the head started really tingling then I immediately started looking for his hand (...) it [was] easier to find his hand with my hand". Moreover, PS4 (and similarly PS3) stated: "I could use hand[-device] to detect more freely, I can use head to detect big thing and used the hand to detect detailed part". PB6 also added to this "[with hands-device] you're working with a smaller space, but with the forehead is a bigger space." Nevertheless, one participant who tried both devices (PS1) used a different strategy from the rest to solve this task: "I ended going up and down with the head[-device] to see where his hand was and where it ended."

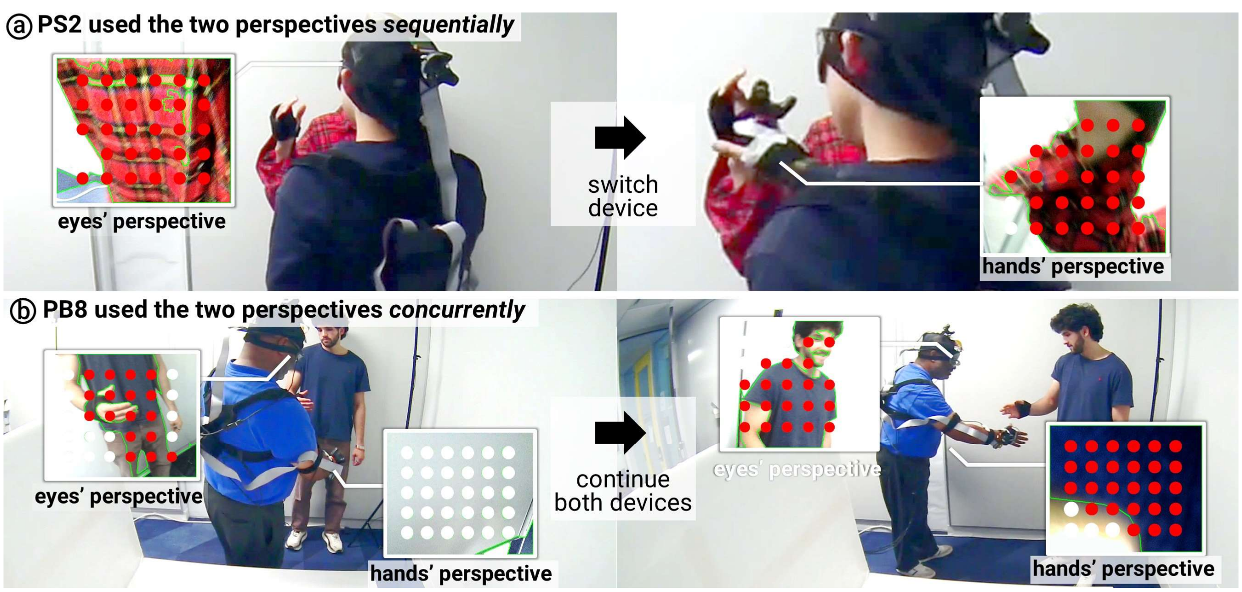

Figure 9: In the final phase of the study participants were offered the option to choose which interface to use to locate a person and shake their hand. (a) PS2 using the two devices sequentially (first, *eye-device* to locate the person, then switched to *hand-device* to locate the hand); (b) PB8 using both devices concurrently for locating the person and hand. (Photos with consent from participants)

When combined, only the *hand-device* moved extensively. While in our previous set of tasks (phase 1, where devices were used in isolation) we observed almost as much movement of the *eye-device* as of the *hand-device*, this was no longer the case when devices were combined (here we relied on video observation since it is not trivial to decouple walking from moving head and hand). During combined usage, only one participant performed significant scanning with the head (PS1), while all others only moved significantly the *hand-device*.

7 ENVISIONED APPLICATIONS FROM OUR STUDY PARTICIPANTS

Finally, we asked our study participants what they would envision using either the combined- or hand-device for. We present their feedback, and additionally, in Figure 10, propose envisioned applications drawing from their experiences.

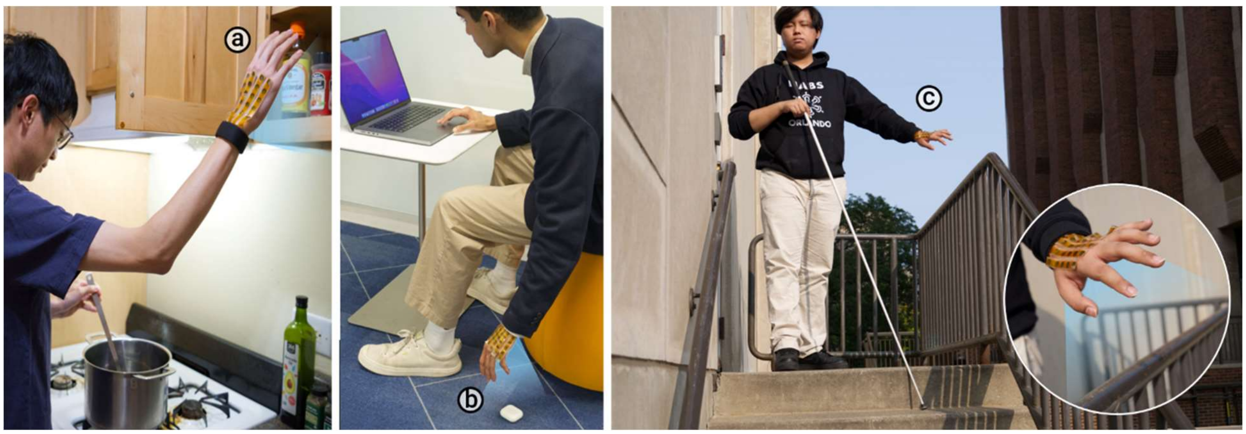

Figure 10: Envisioned applications drawn from participants' feedback proposing using the hands-device to: (a) identify and grab ingredients inside a shelf while cooking (from PS1 & PB6); (b) retrieve an object that fell on the floor, without looking (from PB8 & PB6); (c) find the banister/handrail while walking down the stairs (from PB5).

Situational impairments. Sighted participants envisioned situations where they momentarily could not use their eyes and would rely on the substitutional devices. To this end, PS1 envisioned using the hands-device while cooking with both hands—we depict this envisioned application in Figure 10 (a)—or while walking in the dark and trying to feel where the light switches are (if the camera afforded night vision). PS2 envisioned the hands-device while trying to find objects inside of holes (e.g., pipes or even "in the water"). PS3 envisioned to see "colors" with their hand. Finally, PS4 envisioned the device for "when I'm walking (…) to be more aware of your surroundings".

Blind or Low Vision participants envisioned new combinations (e.g., cane). Our Blind or Low Vision participants also drew extensively from their lived experience as non-visual. For instance, PB5 envisioned a novel use for the combined-device: "the [eyes-]device will (…) know there's something coming, and I will know to move to the left or to the right to navigate around it (…) and the [hand-device] would work well for finding a banister to walk downstairs (…) subway stairs. Because with low vision it's sometimes difficult to find the banister, so me personally I have to hold on to walk down the stairs down"—we depict this envisioned application in Figure 10 (c). PB6 envisioned using the combined-device to "know someone's there and next to you in the airplane or train and I'm not saying it's taking the place of your cane but it's just nice to know". PB6 also envisioned using the hand-device to "reach for things on the shelf, without breaking the glasses"—we depict this envisioned application in Figure 10 (a). PB6 also envisioned using the hands-device for "if I drop earrings on the floor the hand[-device] could be really great for that"—we depict this envisioned application in Figure 10 (b). PB7 envisioned not a specific application but areas in which they could feel that sensory substitution would be advantageous, such as for when working in tight spaces (e.g., for "plumbing or surgery"). Finally, PB8 envisioned many possible use cases for the device, stating: "Earlier today I dropped my cane, and I had to crawl on the floor. But if I had one of these [combined-device] I could follow the [electrotactile]". Then specifically about the hand-device they envisioned: "if I dropped something on the floor (…) under the table I'd try the hand[-device] because if I used the head, I'd have to avoid [hitting] the kitchen table." —we depict this envisioned application in Figure 10 (b).

8 DISCUSSION & LIMITATIONS

Limitations of our study: (1) While our study was already several hours long, it only covered a limited number of tasks. (2) The first phase of our study involved tasks that are not perfectly representative of everyday tasks, since we purposefully restricted users to solving these tasks using sensory substitution devices and did not allow them to solve by touch alone. (3) For simplicity, we only tested unimanual tasks. (4) Our results are limited to the number of recruited participants. (5) The computer vision was limited to contour segmentation, while our tactile resolution was limited by the available channels on our multiplexers (30 electrodes on the hand and 30 on the forehead). Therefore, we warrant caution when generalizing our findings to other contexts, which would require validation from future researchers.

Discussion of our findings with respect to future directions. Importantly, despite the aforementioned study limitations, we see a number of key findings that might illuminate future directions. **(1) Core benefits of our device for manual interactions**: We found that, when compared to traditional eyes' perspective, the hand's perspective afforded more ergonomic ways to identify and grasp objects. Moreover, when given the option to use both devices, we found that participants tended to use the eye-device for *overview* and the hand-device for *details*—some even mentioned this was motivated by the extra mobility of the hands, which made it easier for some participants to focus on a certain region of interest (by moving one's hands instead of needing to move the neck or entire torso/body). These benefits arose even with our fixed setup, one can envision future directions in which users can tweak the camera's angle or field of view (e.g., zoom, tilt, pan)—potentially enabling new flexible interactions for sensory substitution devices. **(2) Combination of multiple "ways of seeing"**. It is worth pausing on the fact that our participants, without much training, were able to use the hand-device, since this interaction is completely novel to the brain—humans only see from their eyes. Being able to switch between two kinds of "seeing" with little training illustrates the flexibility of humans in incorporate new sensory signals [4] (e.g., as it is also the case when learning to use a cane). Based on this result, one might speculate it could be possible to explore parallel perspectives (e.g., "bimanual seeing"), fine-grained perspectives, (e.g., "finger seeing", akin to [49]) or even "seeing through other limbs" (e.g., "seeing with foot"). **(3) Perspective switching UI**: In our study, the participants verbally told the experimenters when they wanted to switch between the substitution devices. It is worth exploring the design of interaction techniques for switching perspectives, e.g., automatically based on inferring attention or intention, or manual using direct manipulation interfaces. **(4) More elaborate tactile-images**: Researchers might want to test new vision algorithms (e.g., depth rather than contours), camera placements (e.g., fingertips), or other visual-tactile mappings beyond binary stimulations (e.g., feeling depth, colors, and so forth). **(5) Impact on spatial understanding**: Participants in our study mentioned they memorized where objects were by scanning with our hand-device. It might be worth investigating how this might impact their spatial mental model. **(6) Social acceptance & privacy**: Lastly, as with any device based on cameras, it may also raise concerns regarding the privacy of users, which is fertile ground for future variations that build on privacy-focused/preserving camera approaches [22].

9 CONCLUSION

Most vision-to-tactile sensory-substitution interfaces focus on translating images captured from the eye's perspective to tactile patterns. We argue that this focus on the eyes' perspective is excellent understanding one's surroundings (e.g., navigation and avoiding obstacles), but misses new interactive opportunities that arise from exploring other vantage points for both the camera and for the tactile output device. As such, we proposed & studied a sensory-substitution device that allows users to *see & feel information from the hands' perspective*. We found that this could enhance flexibility & expressivity of sensory-substitution devices to further support *manual interactions* with physical objects. Our device was engineered as a back-of-the-hand electrotactile-display that renders tactile images from a wrist-mounted camera, allowing the user's hand to feel objects while reaching & hovering. Through our user study with sighted/Blind-or-Low-Vision participants, we found unique benefits of sensory substitution from the hand's perspective-participants felt the hands' perspective was suitable for detailed-oriented work and more ergonomic during object reaching. Nevertheless, in the last phase of the study, we saw how the combination of hands and eyes perspectives was also perceived as beneficial by all participants. We believe these insights extend the landscape of sensory-substitution devices.

ACKNOWLEDGMENTS

We thank Hiroyuki Kajimoto for generously providing the electrotactile toolkit. We are also grateful for the constructive feedback from the reviewers. This work was supported by Google Research Award (Inclusion Research).

REFERENCES

[1] Auvray, M., Hanneton, S. and O'Regan, J.K. 2007. Learning to Perceive with a Visuo — Auditory Substitution System: Localisation and Object Recognition with 'The Voice.' *Perception*. 36, 3 (Mar. 2007), 416–430. https://doi.org/10.1068/p5631.

[2] Bach-y-Rita, P. 2004. Tactile sensory substitution studies. *Annals of the New York Academy of Sciences*. 1013, (May 2004), 83–91. https://doi.org/10.1196/annals.1305.006.

[3] Bach-Y-Rita, P., Collins, C.C., Saunders, F.A., White, B. and Scadden, L. 1969. Vision Substitution by Tactile Image Projection. *Nature*. 221, 5184 (Mar. 1969), 963–964. https://doi.org/10.1038/221963a0.

[4] Bach-y-Rita, P., Tyler, M.E. and Kaczmarek, K.A. 2003. Seeing with the Brain. *International Journal of Human–Computer Interaction*. 15, 2 (Apr. 2003), 285–295. https://doi.org/10.1207/S15327590IJHC1502_6.

[5] Bach-y-Rita, P. and W. Kercel, S. 2003. Sensory substitution and the human–machine interface. *Trends in Cognitive Sciences*. 7, 12 (Dec. 2003), 541–546. https://doi.org/10.1016/j.tics.2003.10.013.

[6] Betti, S., Castiello, U. and Begliomini, C. 2021. Reach-to-Grasp: A Multisensory Experience. *Frontiers in Psychology*. 12, (Feb. 2021). https://doi.org/10.3389/fpsyg.2021.614471.

[7] BrainPort Technologies: *https://www.wicab.com*. Accessed: 2024-08-07.

[8] Brell, M. and Hein, A. 2007. Positioning Tasks in Multimodal Computer-Navigated Surgery. *IEEE MultiMedia*. 14, 4 (Oct. 2007), 42–51. https://doi.org/10.1109/MMUL.2007.81.

[9] Brown, D., Macpherson, T. and Ward, J. 2011. Seeing with Sound? Exploring Different Characteristics of a Visual-to-Auditory Sensory Substitution Device. *Perception*. 40, 9 (Sep. 2011), 1120–1135. https://doi.org/10.1068/p6952.

[10] Chebat, D.-R., Schneider, F.C., Kupers, R. and Ptito, M. 2011. Navigation with a sensory substitution device in congenitally blind individuals. *NeuroReport*. 22, 7 (May 2011), 342. https://doi.org/10.1097/WNR.0b013e3283462def.

[11] Collins, C.C. and Bach-y-Rita, P. 1973. Transmission of Pictorial Information Through the Skin. *Advances in Biological and Medical Physics*. 14, (Jan. 1973), 285–315. https://doi.org/10.1016/B978-0-12-005214-1.50010-8.

[12] Cronly-Dillon, J., Persaud, K.C. and Blore, R. 2000. Blind subjects construct conscious mental images of visual scenes encoded in musical form. *Proceedings. Biological Sciences*. 267, 1458 (Nov. 2000), 2231–2238. https://doi.org/10.1098/rspb.2000.1273.

[13] De Paz, C., Ibáñez-Gijón, J., Travieso, D. and Jacobs, D.M. 2023. Grasping objects with a sensory substitution glove. *International Journal of Human-Computer Studies*. 170, (Feb. 2023), 102963. https://doi.org/10.1016/j.ijhcs.2022.102963.

[14] Delleman, N.J. and Dul, J. 2007. International standards on working postures and movements ISO 11226 and EN 1005-4. *Ergonomics*. 50, 11 (Nov. 2007), 1809–1819. https://doi.org/10.1080/00140130701674430.

[15] Eagleman, D.M. and Perrotta, M.V. 2023. The future of sensory substitution, addition, and expansion via haptic devices. *Frontiers in Human Neuroscience*. 16, (Jan. 2023). https://doi.org/10.3389/fnhum.2022.1055546.

[16] Geed, S. and van Kan, P.L.E. 2017. Grasp-Based Functional Coupling Between Reach- and Grasp-Related Components of Forelimb Muscle Activity. *Journal of motor behavior*. 49, 3 (2017), 312–328. https://doi.org/10.1080/00222895.2016.1204265.

[17] Germani, M., Mengoni, M. and Peruzzini, M. 2013. Electro-tactile device for material texture simulation. *The International Journal of Advanced Manufacturing Technology*. 68, 9 (Oct. 2013), 2185–2203. https://doi.org/10.1007/s00170-013-4832-1.

[18] Goodale, M.A. 2011. Transforming vision into action. *Vision Research*. 51, 13 (Jul. 2011), 1567–1587. https://doi.org/10.1016/j.visres.2010.07.027.

[19] Günther, S., Müller, F., Funk, M., Kirchner, J., Dezfuli, N. and Mühlhäuser, M. 2018. TactileGlove: Assistive Spatial Guidance in 3D Space through Vibrotactile Navigation. *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference* (Corfu Greece, Jun. 2018), 273–280. https://doi.org/10.1145/3197768.3197785.

[20] Hart, S.G. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 50, 9 (Oct. 2006), 904–908. https://doi.org/10.1177/154193120605000909.

[21] Horvath, S., Galeotti, J., Wu, B., Klatzky, R., Siegel, M. and Stetten, G. 2014. FingerSight: Fingertip Haptic Sensing of the Visual Environment. *IEEE Journal of Translational Engineering in Health and Medicine*. 2, (2014), 1–9. https://doi.org/10.1109/JTEHM.2014.2309343.

[22] Iravantchi, Y., Krolikowski, T., Wang, W., Shin, K.G. and Sample, A. 2024. PrivacyLens: On-Device PII Removal from RGB Images using Thermally-Enhanced Sensing. *Proceedings on Privacy Enhancing Technologies*. (2024).

[23] Israr, A. and Poupyrev, I. 2011. Tactile brush: drawing on skin with a tactile grid display. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2011), 2019–2028. https://doi.org/10.1145/1978942.1979235.

[24] Jakobson, L.S. and Goodale, M.A. 1991. Factors affecting higher-order movement planning: a kinematic analysis of human prehension. *Experimental Brain Research*. 86, 1 (Aug. 1991), 199–208. https://doi.org/10.1007/BF00231054.

[25] Jingu, A., Withana, A. and Steimle, J. 2023. Double-Sided Tactile Interactions for Grasping in Virtual Reality. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, Oct. 2023), 1–11. https://doi.org/10.1145/3586183.3606798.

[26] Jung, J., Son, S., Lee, S., Kim, Y. and Lee, G. 2021. ThroughHand: 2D Tactile Interaction to Simultaneously Recognize and Touch Multiple Objects. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2021), 1–13. https://doi.org/10.1145/3411764.3445530.

[27] Kaczmarek, K.A. 2011. The tongue display unit (TDU) for electrotactile spatiotemporal pattern presentation. *Scientia Iranica*. 18, 6 (Dec. 2011), 1476–1485. https://doi.org/10.1016/j.scient.2011.08.020.

[28] Kaczmarek, K.A., Webster, J.G., Bach-y-Rita, P. and Tompkins, W.J. 1991. Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomedical Engineering*. 38, 1 (Jan. 1991), 1–16. https://doi.org/10.1109/10.68204.

[29] Kajimoto, H. 2021. Electro-Tactile Display Kit for Fingertip. *2021 IEEE World Haptics Conference (WHC)* (Jul. 2021), 587–587. https://doi.org/10.1109/WHC49131.2021.9517192.

[30] Kajimoto, H. 2016. Electro-tactile Display: Principle and Hardware. *Pervasive Haptics: Science, Design, and Application*. H. Kajimoto, S. Saga, and M. Konyo, eds. Springer Japan. 79–96. https://doi.org/10.1007/978-4-431-55772-2_5.

[31] Kajimoto, H., Kannno, Y. and Tachi, S. Forehead Electro-tactile Display for Vision Substitution. *EuroHaptics 2006*.

[32] Katzschmann, R.K., Araki, B. and Rus, D. 2018. Safe Local Navigation for Visually Impaired Users With a Time-of-Flight and Haptic Feedback Device. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 26, 3 (Mar. 2018), 583–593. https://doi.org/10.1109/TNSRE.2018.2800665.

[33] Kilian, J., Neugebauer, A., Scherffig, L. and Wahl, S. 2022. The Unfolding Space Glove: A Wearable Spatio-Visual to Haptic Sensory Substitution Device for Blind People. *Sensors*. 22, 5 (Jan. 2022), 1859. https://doi.org/10.3390/s22051859.

[34] Kourtesis, P., Argelaguet, F., Vizcay, S., Marchal, M. and Pacchierotti, C. 2021. Electrotactile feedback for hand interactions:A systematic review, meta-analysis,and future directions. *arXiv:2105.05343 [cs]*. (May 2021).

[35] Krishna, S., Bala, S. and Panchanathan, S. 2010. Exploring the dorsal surface of the fingers for visuo-haptic sensory substitution. *2010 IEEE International Symposium on Haptic Audio Visual Environments and Games* (Phoenix, AZ, USA, Oct. 2010), 1–6. https://doi.org/10.1109/HAVE.2010.5623974.

[36] Kristjánsson, Á., Moldoveanu, A., Jóhannesson, Ó.I., Balan, O., Spagnol, S., Valgeirsdóttir, V.V. and Unnthorsson, R. Designing sensory-substitution devices: Principles, pitfalls and potential. *Restorative Neurology and Neuroscience*. 34, 5, 769–787. https://doi.org/10.3233/RNN-160647.

[37] Land, M., Mennie, N. and Rusted, J. 1999. The Roles of Vision and Eye Movements in the Control of Activities of Daily Living. *Perception*. 28, 11 (Nov. 1999), 1311–1328. https://doi.org/10.1068/p2935.

[38] Lee, J., Tjahjadi, A.D., Kim, J., Yu, J., Park, M., Zhang, J., Froehlich, J.E., Tian, Y. and Zhao, Y. 2024. CookAR: Affordance Augmentations in Wearable AR to Support Kitchen Tool Interactions for People with Low Vision. *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, Oct. 2024), 1–16. https://doi.org/10.1145/3654777.3676449.

[39] Lewis, S. 2015. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. *Health Promotion Practice*. 16, 4 (Jul. 2015), 473–475. https://doi.org/10.1177/1524839915580941.

[40] Lloyd-Esenkaya, T., Lloyd-Esenkaya, V., O'Neill, E. and Proulx, M.J. 2020. Multisensory inclusive design with sensory substitution. *Cognitive Research: Principles and Implications*. 5, 1 (Aug. 2020), 37. https://doi.org/10.1186/s41235-020-00240-7.

[41] Lobo, L., Travieso, D., Barrientos, A. and Jacobs, D.M. 2014. Stepping on Obstacles with a Sensory Substitution Device on the Lower Leg: Practice without Vision Is More Beneficial than Practice with Vision. *PLOS ONE*. 9, 6 (Jun. 2014), e98801. https://doi.org/10.1371/journal.pone.0098801.

[42] Lobo, L., Travieso, D., Jacobs, D.M., Rodger, M. and Craig, C.M. 2018. Sensory substitution: Using a vibrotactile device to orient and walk to targets. *Journal of Experimental Psychology: Applied*. 24, 1 (2018), 108–124. https://doi.org/10.1037/xap0000154.

[43] Mateevitsi, V., Haggadone, B., Leigh, J., Kunzer, B. and Kenyon, R.V. 2013. Sensing the environment through SpiderSense. *Proceedings of the 4th Augmented Human International Conference* (New York, NY, USA, Mar. 2013), 51–57. https://doi.org/10.1145/2459236.2459246.

[44] Meijer, P.B.L. 1992. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*. 39, 2 (Feb. 1992), 112–121. https://doi.org/10.1109/10.121642.

[45] Mon-Williams, M., Tresilian, J.R., Coppard, V.L. and Carson, R.G. 2001. The effect of obstacle position on reach-to-grasp movements. *Experimental Brain Research*. 137, 3 (Apr. 2001), 497–501. https://doi.org/10.1007/s002210100684.

[46] OpenCV: Canny Edge Detection: *https://docs.opencv.org/4.x/da/d22/tutorial_py_canny.html*. Accessed: 2024-09-10.

[47] OpenCV: Contour Features: *https://docs.opencv.org/4.x/dd/d49/tutorial_py_contour_features.html*. Accessed: 2024-09-10.

[48] Powell, P., Pätzold, F., Rouygari, M., Furtak, M., Kärcher, S.M. and König, P. 2024. Helping Blind People Grasp: Evaluating a Tactile Bracelet for Remotely Guiding Grasping Movements. *Sensors*. 24, 9 (Jan. 2024), 2949. https://doi.org/10.3390/s24092949.

[49] Satpute, S.A., Canady, J.R., Klatzky, R.L. and Stetten, G.D. 2020. FingerSight: A Vibrotactile Wearable Ring for Assistance With Locating and Reaching Objects in Peripersonal Space. *IEEE Transactions on Haptics*. 13, 2 (Apr. 2020), 325–333. https://doi.org/10.1109/TOH.2019.2945561.

[50] Schlereth, T., Magerl, W. and Treede, R.-D. 2001. Spatial discrimination thresholds for pain and touch in human hairy skin. *PAIN*. 92, 1 (May 2001), 187. https://doi.org/10.1016/S0304-3959(00)00484-X.

[51] Shin, B.-S. and Lim, C.-S. 2007. Obstacle Detection and Avoidance System for Visually Impaired People. *Haptic and Audio Interaction Design* (Berlin, Heidelberg, 2007), 78–85. https://doi.org/10.1007/978-3-540-76702-2_9.

[52] Shin, K., Lee, M., Chang, M., Bae, Y.M., Chang, W. and Kim, Y.-J. 2024. Enhancing Visual Perception for People with Blindness: A Feasibility Study of a 12-Channel Forehead ElectroTactile Stimulator with a Stereo Camera.

[53] Stanke, D., Duente, T. and Rohs, M. 2020. TactileWear: A Comparison of Electrotactile and Vibrotactile Feedback on the Wrist and Ring Finger. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (New York, NY, USA, Oct. 2020), 1–13. https://doi.org/10.1145/3419249.3420107.

[54] Stephens-Fripp, B., Alici, G. and Mutlu, R. 2018. A Review of Non-Invasive Sensory Feedback Methods for Transradial Prosthetic Hands. *IEEE Access*. 6, (2018), 6878–6899. https://doi.org/10.1109/ACCESS.2018.2791583.

[55] Strong, R.M. and Troxel, D.E. 1970. An Electrotactile Display. *IEEE Transactions on Man-Machine Systems*. 11, 1 (Mar. 1970), 72–79. https://doi.org/10.1109/TMMS.1970.299965.

[56] Tan, C., Tretriluxana, J., Pitsch, E., Runnarong, N. and Winstein, C.J. 2012. Anticipatory Planning of Functional Reach-to-Grasp: A Pilot Study. *Neurorehabilitation and neural repair*. 26, 8 (Oct. 2012), 957–967. https://doi.org/10.1177/1545968312437938.

[57] Tanaka, Y., Shen, A., Kong, A. and Lopes, P. 2023. Full-hand Electro-Tactile Feedback without Obstructing Palmar Side of Hand. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, Apr. 2023), 1–15. https://doi.org/10.1145/3544548.3581382.

[58] Tashiro, T. and Higashiyama, A. 1981. The perceptual properties of electrocutaneous stimulation: Sensory quality, subjective intensity, and intensity-duration relation. *Perception & Psychophysics*. 30, 6 (Nov. 1981), 579–586. https://doi.org/10.3758/BF03202013.

[59] Teng, S.-Y., Gupta, A. and Lopes, P. 2024. Haptic Permeability: Adding Holes to Tactile Devices Improves Dexterity. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2024), 1–12. https://doi.org/10.1145/3613904.3642156.

[60] The vOICe - New Frontiers in Artificial Vision: *https://www.seeingwithsound.com/*. Accessed: 2024-09-11.

[61] Ushiyama, K. and Lopes, P. 2023. FeetThrough: Electrotactile Foot Interface that Preserves Real-World Sensations. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco CA USA, Oct. 2023), 1–11. https://doi.org/10.1145/3586183.3606808.

[62] Visell, Y. 2009. Tactile sensory substitution: Models for enaction in HCI. *Interacting with Computers*. 21, 1 (Jan. 2009), 38–53. https://doi.org/10.1016/j.intcom.2008.08.004.

[63] Weber, B., Schätzle, S., Hulin, T., Preusche, C. and Deml, B. 2011. Evaluation of a vibrotactile feedback device for spatial guidance. *World Haptics Conference (WHC), 2011 IEEE* (2011), 349–354. https://doi.org/10.1109/WHC.2011.5945511.

[64] Wing, A.M., Turton, A. and Fraser, C. 1986. Grasp Size and Accuracy of Approach in Reaching. *Journal of Motor Behavior*. 18, 3 (Sep. 1986), 245–260. https://doi.org/10.1080/00222895.1986.10735380.

[65] Winges, S.A., Weber, D.J. and Santello, M. 2003. The role of vision on hand preshaping during reach to grasp. *Experimental Brain Research*. 152, 4 (Oct. 2003), 489–498. https://doi.org/10.1007/s00221-003-1571-9.

[66] Withana, A., Groeger, D. and Steimle, J. 2018. Tacttoo: A Thin and Feel-Through Tattoo for On-Skin Tactile Output. *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, Oct. 2018), 365–378. https://doi.org/10.1145/3242587.3242645.

[67] Yem, V. and Kajimoto, H. 2017. Wearable tactile device using mechanical and electrical stimulation for fingertip interaction with virtual world. *2017 IEEE Virtual Reality (VR)* (Mar. 2017), 99–104. https://doi.org/10.1109/VR.2017.7892236.

[68] Yu, Z., Horvath, S., Delazio, A., Wang, J., Almasi, R., Klatzky, R., Galeotti, J. and Stetten, G. PalmSight: an assistive technology helping the blind to locate and grasp objects.