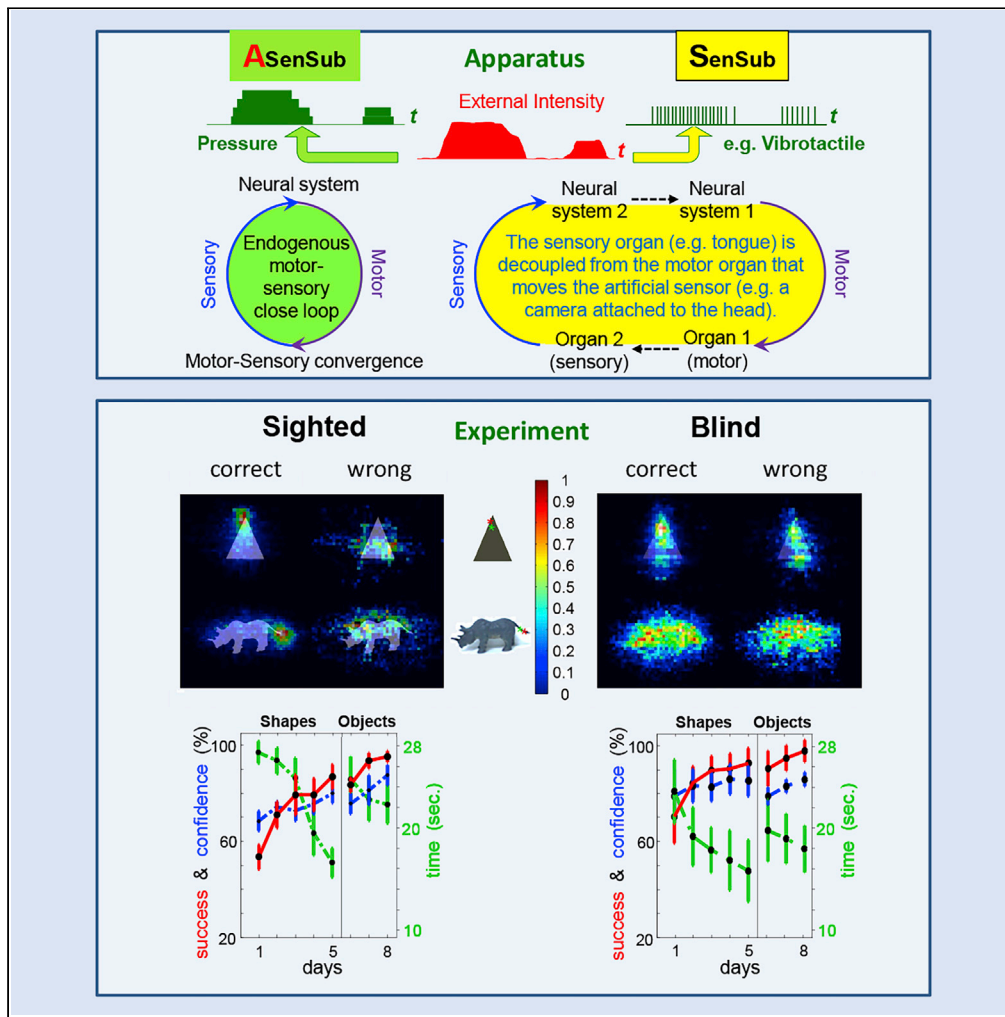


Article

Active sensory substitution allows fast learning via effective motor-sensory strategies



Yael Zilbershtain-Kra, Shmuel Graffi, Ehud Ahissar, Amos Arieli

ehud.ahissar@weizmann.ac.il (E.A.)
amos.arieli@weizmann.ac.il (A.A.)

Highlights

We describe sensory substitution based on active-sensing principles (ASenSub)

Perception via ASenSub emerges rapidly and transferred easily from 2D to 3D objects

Sighted and blind users employ visual- and tactile-like strategies, respectively

Use of natural motor-sensory strategies facilitates learning and performance



Article

Active sensory substitution allows fast learning via effective motor-sensory strategies

Yael Zilbershtain-Kra,¹ Shmuel Graffi,¹ Ehud Ahissar,^{1,2,*} and Amos Arieli^{1,2,3,*}

Summary

We examined the development of new sensing abilities in adults by training participants to perceive remote objects through their fingers. Using an Active-Sensing based sensory Substitution device (ASenSub), participants quickly learned to perceive fast via the new modality and preserved their high performance for more than 20 months. Both sighted and blind participants exhibited almost complete transfer of performance from 2D images to novel 3D physical objects. Perceptual accuracy and speed using the ASenSub were, on average, 300% and 600% better than previous reports for 2D images and 3D objects. This improvement is attributed to the ability of the participants to employ their own motor-sensory strategies. Sighted participants dominant strategy was based on motor-sensory convergence on the most informative regions of objects, similarly to fixation patterns in vision. Congenitally, blind participants did not show such a tendency, and many of their exploratory procedures resembled those observed with natural touch.

Introduction

Vision and touch share a common sensory strategy: Both modalities are based on two-dimensional (2D) arrays of receptors that actively scan the environment (Pei et al., 2010; Ahissar and Arieli, 2001; Collins and Bach-Y-Rita, 1973; Borji et al., 2013). Humans, and primates in general, typically move their eyes and hands while perceiving external objects, in patterns whose interactions with the objects determine receptor activations (Ahissar and Vaadia, 1990; Connor and Johnson, 1992; Kelso, 1997; Gamzu and Ahissar, 2001; Visell, 2009; Yoshioka et al., 2011; Ahissar and Arieli, 2012; Mannan et al., 2009). In both systems, this active-sensing strategy is supported by an elaborated system of parallel, multi-level motor-sensory-motor loops (Ahissar and Assa, 2016) and induces spatiotemporal sensory coding (Ahissar and Arieli, 2001, 2012; Pei et al., 2010; Saig et al., 2012; Ahissar and Vaadia, 1990; Gamzu and Ahissar, 2001; Hollins and Bensmaïa, 2007; Cascio and Sathian, 2001; Sathian, 1989). We examined motor-sensory exploration strategies underlying the development of novel sensing abilities by training participants to perceive through their fingers remote objects, typically perceived via vision—a method known as vision-to-touch sensory substitution system (SenSub) (Bach-Y-Rita et al., 1969).

SenSub allows the acquisition of environmental information that is usually acquired through one sense by another sense (Lenay et al., 2003; Zilbershtain-Kra et al., 2015; Collins and Bach-Y-Rita, 1973). Visual to tactile SenSub allows people who have impaired vision to see aspects of their environment via tactile signals. While most existing vision-to-touch SenSubs preserve the 2D sensory acquisition, they abandon the active-sensing component: The user does not move the organ that receives the tactile stimulation (e.g. finger, tongue), and the SenSub typically converts steady visual inputs into skin vibrations (applied to prevent receptor adaptation). Perception via such SenSub emerges only after long training periods (Arno et al., 1999; Maidenbaum et al., 2013; Striem-Amit et al., 2012; Ziat et al., 2005) and often remains frustratingly slow even after training (Nye and Bliss, 1970; Amedi et al., 2007). Perception becomes easier when participants are allowed to move the artificial sensor (e.g., camera) with their hands (Lenay et al., 2003; Doel et al., 2004; Auvray et al., 2007; Brown et al., 2011; Pfitz et al., 2005). Only in one study, as far as we know, the motion-sensation loop was closed in order to mimic active touch. In that study, the participants placed their fingers onto a carriage which could be moved over the surface of the device to reveal a virtual shape (Chan

¹The Department of Neurobiology, Weizmann Institute of Science, 234 Herzl Street, Rehovot 76100, Israel

²These authors contributed equally

³Lead contact

*Correspondence: ehud.ahissar@weizmann.ac.il (E.A.), amos.arieli@weizmann.ac.il (A.A.)

<https://doi.org/10.1016/j.isci.2020.101918>



et al., 2007). Such mimicry of active touch allowed good performance with little training (see Discussion and Table S1).

Although SenSubs exist for more than 50 years, none of them has been adapted so far by the blind community for practical everyday use. In this study, we assumed that the primary barrier is these devices' incompatibility with natural perceptual strategies and their failure to allow the use of common sense mechanisms. Thus, we constructed a SenSub that is based on two major principles of natural active-sensing: the sensory (tactile) organ actively acquires its (visual) information, and no artificial aid is used to prevent receptor adaptation. Our SenSub allowed, and in fact forced, active-sensing: participants had to move their hands in order to sense steady objects, much as they would do in order to palpate a proximal object via touch or to see it with their eyes, and sensation occurred at the moving hand.

As this active-sensing based SenSub device (ASenSub) introduces an unfamiliar sensory modality to naive participants, it allowed the tracking of the principles underlying the development and the use of perceptual strategies in adults. We show here that practicing with our ASenSub quickly leads to accurate and rapid perception, which is based primarily on focused scanning of an informative region of the object, similar to a fixation pattern observed with natural vision (Reinagel and Zador, 1999; Schumann et al., 2008; Parkhurst and Niebur, 2003).

Results

We used a large prototype of ASenSub, which allowed good stimulus control and high-resolution tracking. The tactile actuator was attached to the hand of each participant such that three fingers were constantly contacting three touch arrays of 32 pins each, with fixed inter-finger distances (Figure 1, #1), resulting in an array of 96 (12*8) pins across the three fingers. The array of pins was mapped to a fixed array of 192*128 pixels at the center of the video camera that was attached to the same hand (Figure 1, #2). The height of each pin was directly proportional to the mean illumination level captured by its corresponding array of 16*16 video pixels. The external information was thus constantly given via the ASenSub by simply translating external contrasts to skin indentations—no further processing was added. Accordingly, steady illumination levels yielded steady deflection levels (without any vibrations) and hand movements induced spatiotemporal patterns of skin deflections similar to those induced during natural tactile scanning of an object (Connor and Johnson, 1992; Weber et al., 2013; Mackevicius et al., 2012).

The use of our ASenSub was tested in six experiments of five alternative forced-choice identification task, with two groups of six participants each. In each experiment, a different set of objects was used, including two sets of 2D computer images that were presented on a computer screen, and one set of 3D physical objects that were presented on a table (Figure 2A). The participants were free to move their hands (with the ASenSub) in three dimensions (Figure 1, #4) thus also to change the distance between the ASenSub and the object. No change was required in the setup when moving from exploring 2D to exploring 3D objects, except for replacing the objects. The perception of 20 objects was tested in each session. Hand movements, pin deflections, participants' on-going introspective audio-reports, self-confidence levels and decisions were registered for every trial.

Rapid learning, high-performance, and long-term retention of blindfolded sighted participants

In the first experiment (EXP 1), sighted participants ($n = 6$) showed a rapid improvement in performance, reaching $86.7 \pm 5.4\%$ on the fifth training sessions, each containing 20 trials. Self-confidence levels increased and recognition times decreased in parallel to the increase in success rate, reaching 16.4 ± 1.6 s on the fifth session (Figure 2B, 'EXP 1'). EXP 2, which immediately followed EXP 1, was similar to EXP 1 but with 3D objects (Figure 2A, second column). Remarkably, even though the 3D objects had nothing in common with the 2D images, performance levels were transferred almost completely from EXP 1 to the first session of EXP 2 (Figure 2B, day 6).

Moreover, during the next two experimental sessions (7 and 8), participants further improved, reaching a success rate of $95 \pm 2.6\%$ on the eighth session (Video S1), with their confidence further increasing (Figure 2B, "EXP 2"). Acquisition times in EXP 2 started at the level reached for the same trial duration (30 s) in EXP 1 (session 3) and slightly decreased upon training (to 22.2 ± 2.1 s). In EXPs 1 and 2 the participants'

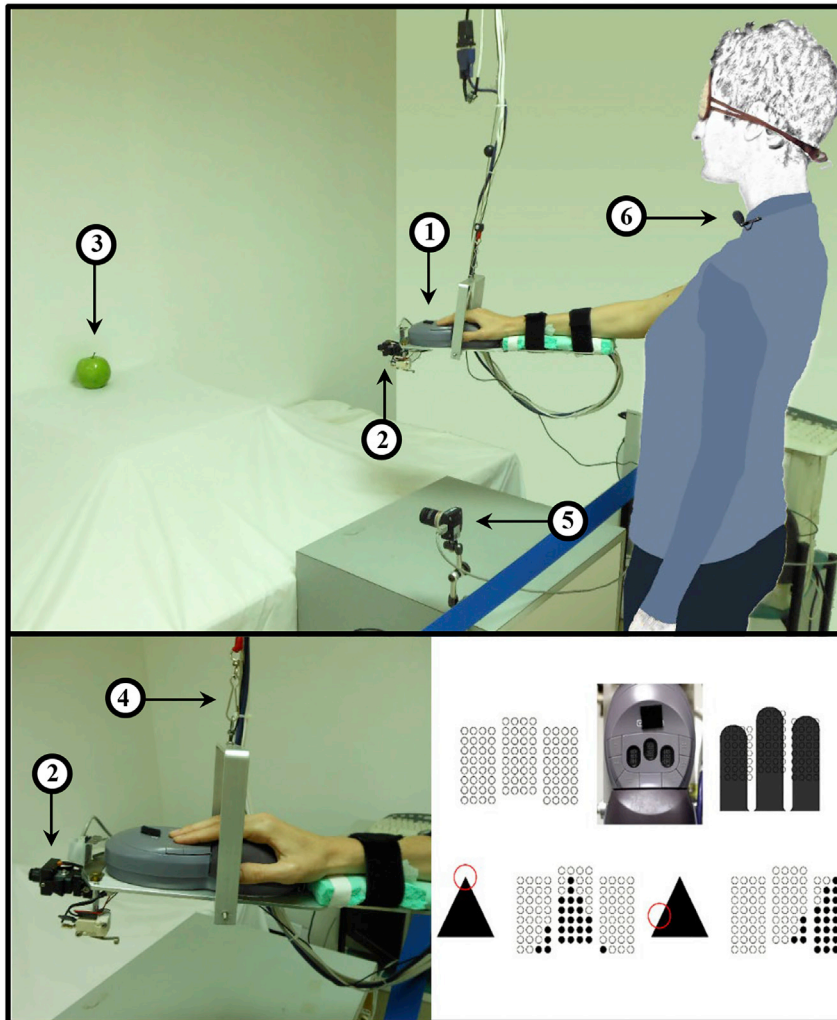


Figure 1. Experimental setup

A picture of the experimental setup with schematic illustration of a blindfolded participant while identifying an apple (#3). The tactile actuator (#1)—array of pins stimulating by variable pressure—is firmly attached to the participant’s hand such that the three middle fingers are constantly contacting three touch arrays of 32 pins each (4*8). The height of each pin (4 levels) is directly related to the illumination level captured by a corresponding pixel array in the active camera (#2). A wire from a balancer (#4) attached to the ceiling takes part of the system’s weight, while enabling effortless 3D hand movement. A fixed wide-field video camera (#5) is stationed, in order to picture the whole of the experimental scene. Offline analysis of the video flow from the two cameras enabled to track the participant’s path, in absolute coordinates, including the object. A small microphone is attached to the participant’s chest (#6). Bottom-Left: Zoom into the ASenSub with the active camera [VQ25B-P37P; Filtech Corp.; Yangchon-Gu Seoul, Korea, frame rate of 28 Hz]. Bottom-Right: Picture of the tactile mouse [VTMouse, Tactile World, refresh rate of 10Hz], and schematic illustration of the pin array. The figure show example while identifying image of triangle (black pins are up). The red circle denote the image captured by the active camera which is transferred to the actuator.

confidence level was highly correlated with their success rate both across objects ($r = 0.86$, $p = 0.001$) and participants ($r = 0.72$, $p = 0.02$). Additionally, confidence level was significantly higher in the correct trials ($p < 0.0001$; 2-sample t test).

Furthermore, participants’ performance levels were preserved for a long period: When tested after >20 months on the same shapes as in EXP 1, the participants ($n = 4$ of the original participants) again started with the level obtained at the end of EXP 1 and rapidly further improved (Figure 2B, “EXP 3”). The same participants were then tested with novel shapes that they did not see before, shapes whose embossed

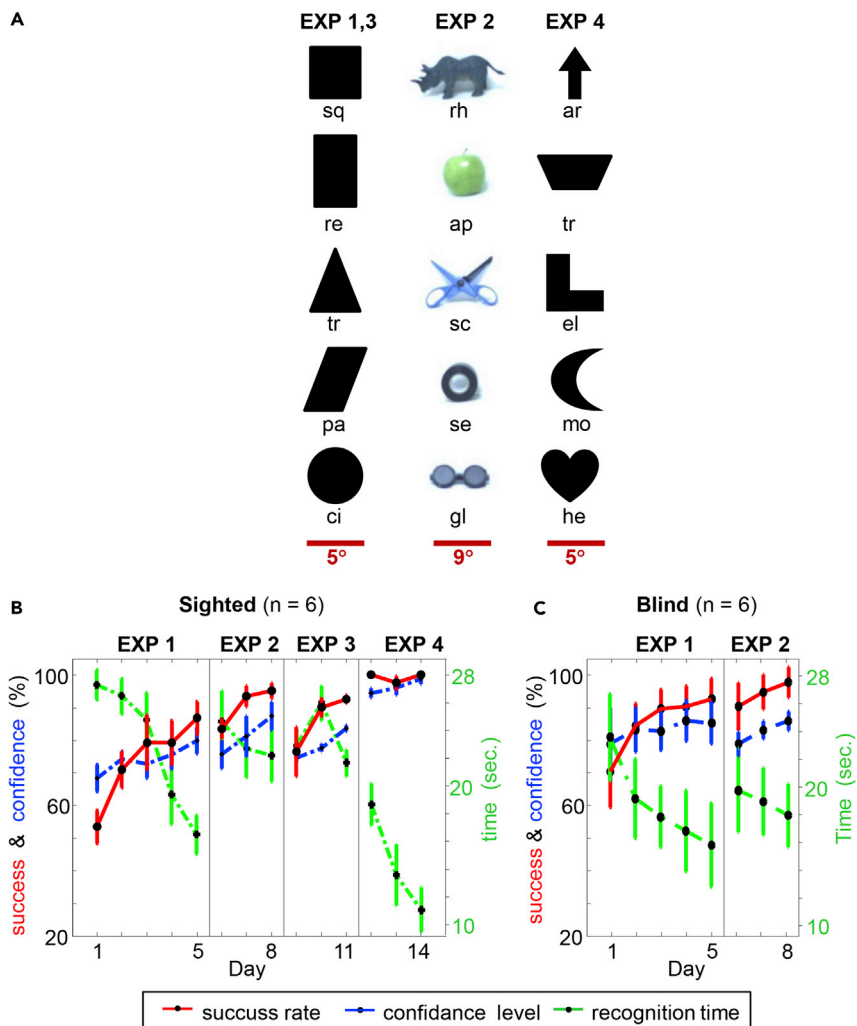


Figure 2. Stimuli and performance

(A) The 3 sets of 5 stimuli used in the different experiments. The set of EXP 2 contained 3-dimensional physical objects and the two other sets contained 2-dimensional shapes presented on a computer screen.

(B) Learning curves showing success rate (solid red), self-confidence level (dashed blue) and identification time (dashed green) of sighted participants (means \pm SEM across participants; n = 6 in EXP 1–2 and n = 4 in EXP 3–4).

(C) EXP1-2 [same as in (B)] with blind participants (n = 6).

images were tactually palpated by the participants before the experiment began (Figure 2B, “EXP 4”; see Methods). With these unseen shapes, their mean success rate reached a ceiling of 100% and recognition time decreased to exceptionally short values (10.9 ± 1.9 s; Figure 2B, “EXP 4”), with the fastest participant identifying shapes in 6.9 ± 0.7 s.

Rapid learning and high performances of congenitally blind participants

Although the ASenSub allowed fast learning and fast and accurate perception with sighted blindfolded participants, it was not at all clear that it would be similarly advantageous for blind participants, and especially so for congenitally blind ones, who cannot rely on any visual priors. We thus trained a group of 6 congenitally blind participants with our ASenSub, using the same protocol used with the blindfolded participants. The participants were allowed to palpate objects having the same profile as the images for a few seconds before the first session. Using the ASenSub, these participants showed fast learning curves and high success rates, similar to our sighted participants (Figure 2C, “EXP 1”). Again, when moving to new 3D objects (Figure 2C, “EXP 2”), performance levels were transferred and continued to increase, while perceptual duration continued to decrease.

Object scanning strategy: focal versus spread scanning

Thus, with short training of several hours with our ASenSub, blind and blindfolded participants could perceive 2D shapes and 3D objects fast and accurately. The quick transfer between 2D to 3D perception could reflect an efficient learning of appropriate exploration strategies. In order to elucidate the motor strategies underlying this fast active perception, we analyzed the scanning trajectories of our participants' hands during the tasks. In general, our participants exhibited a wealth of scanning patterns (see also (Ziat et al., 2005; Chan et al., 2007)), whose details and general characteristics changed from trial to trial and session to session. Despite this large variability, several consistent components of scanning were evident: the hands scanned across or along contours, and often dwelled on certain regions of the object.

We classified motion segments according to their relationships with the shapes or object contour, with the major distinction being between motions that followed the contours and those that crossed it (Figure 3A). Overall, both sighted and blind participants exhibit mixed patterns of scanning, with a general common rule—the higher the proportion of contour following was in a trial, the slower the mean scanning speed was (Figure 3B, $r = -0.93$, $p < 0.0001$).

Sighted and blind participant, however, did differ in their tendency to dwell in salient object features. The common scanning patterns of each participant-object pair were depicted by the grand averages of the visit rate at each pixel of the object's cross section. The visit rates of the sighted participants, in particular in correct trials, demonstrated a pattern of focal scanning of an object-specific region (focal region, or FocR) (Figure 4A left); the participants scanned the FocR more frequently and lingered there more than in other regions. For each object, the FocR was defined as the region around the primary Focal Point (FocP, Figure 5B red asterisks, see Transparent Methods). Quantifying the tendency to dwell in a focal point by the FocPdistance (the mean distance from the FocP during the trial; see STAR Methods) revealed a significant difference between correct and wrong trials (Figure 4B left; 86.7 ± 2.8 vs. 102.3 ± 2.8 pixels; $p < 0.01$, paired t test).

In contrast to the sighted participants, focal scanning was not a prominent strategy of our blind participants (Figure 4A right). This is indicated by their large FocPdistance in the correct trials relative to that of the sighted participants (Figure 4B), and the lack of a consistent difference between correct and wrong trials (Figure 4B right; sighted: $p < 0.01$; blind: $p = 0.6$; paired t-test).

Convergence on objects' informative regions

This strategy of focal scanning was learned by experience, as shown by the gradual decrease in the mean distance of the scanning trajectory from the FocP in parallel to learning (Figure 5A, magenta and red, respectively; EXP 1: $r = -0.98$, $n = 5$ sessions, $p < 0.001$, 1-way ANOVA). Moreover, with learning, perceptual errors became gradually more dependent on this strategy (Figure 5A, cyan curves; asterisks denote significant effects).

To examine the dynamics of scanning in relation to the FocR we quantified a "convergence index" as the percentage of time spent within the FocR during the second half of the trial minus that during the first half (see Transparent Methods). For most of the objects, the convergence index was positive (Figure 5C, abscissa), that is, participants spent more time in scanning the FocR during the second half of the trial. Across objects the convergence index was correlated with the mean success rate such that objects associated with more success were also associated with stronger convergence on the FocR (Figure 5C; $r = 0.86$, $p = 0.001$).

In order to investigate the possible sources of the large differences in convergence indices between objects, we analyzed the structure of spatial information in each object by calculating entropy maps, using entropy as a measure of the amount of spatial information (see Transparent Methods). Conjecturing that our participants attempted to focus on regions with high entropy, we identified the sites within each object that showed the highest entropy (entropy peak, or EP). Objects differed in the number of EPs they exhibited (Figure 5B, green asterisks): five objects (tr, rh, sc, gl, ap) exhibited a single EP and were termed "single-EP objects" and five exhibited multiple EPs and termed "multiple-EP objects". Crucially, the FocPs of our participants overlapped with the EPs of the single-EP objects in a striking manner, but not with those of the multiple-EP objects (Figure 5B). The mean distance between the FocPs and the nearest EPs was an excellent predictor of the success rates across objects (Figure 5D; $r = -0.816$, $p = 0.003$).

Consistently, the sighted participants exhibited different identification dynamics when scanning single- and multiple-EP objects. While scanning single-EP objects, the participants were on average closer to the FocP during

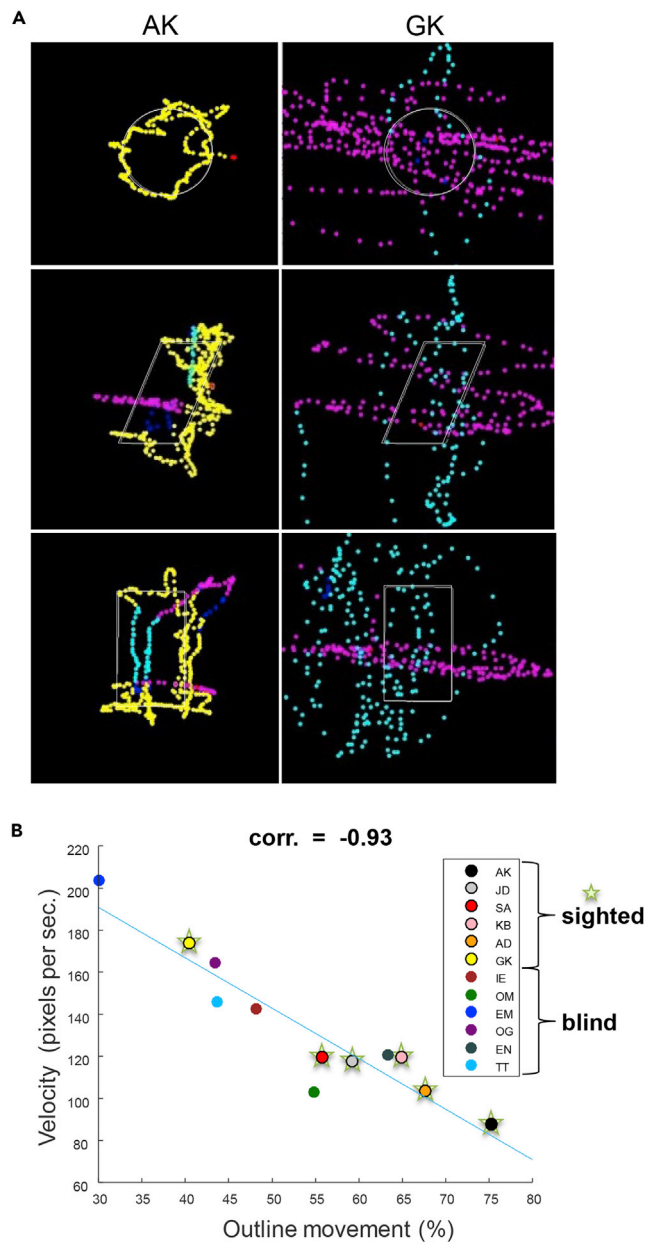


Figure 3. Scanning patterns in individual trials

(A) Single trials from the first session of two participants (AK, 76% correct in session 1, and GK, 50% correct in session 1) scanning (top to bottom) a circle, a rectangle and a parallelogram. Every dot represents hand position at a single video frame overlaid on the object's cross section (only the contour of the object is shown, white). The dots are colored according to their scanning context as follows: Yellow, contour following; magenta, horizontal scanning; cyan, vertical scanning; blue, undefined. In two trials (uppermost left), the 5 first frames are colored red.

(B) The averaged scanning speed for each participant for all shapes as a function of the percent of trial time spent on contour following. Star shapes denote the sighted participants.

the second half of each correct trial (Figure 5F, solid green line), and spent relatively more time there (Figure 5G, solid green line). These dynamics were not exhibited during wrong trials (Figures 5F and 5G, dashed green lines) or while scanning multiple-EP objects (Figures 5F and 5G, black lines). Convergence on FocP evolved along the trials when palpating single-EP objects; the distance from the FocP in single-EP objects gradually decreased during the last several seconds of each trial (Figure 5H, green curve; see Transparent Methods). Such convergence dynamics was not exhibited with multiple-EP objects (Figure 5H, black curve). Discrimination of different objects

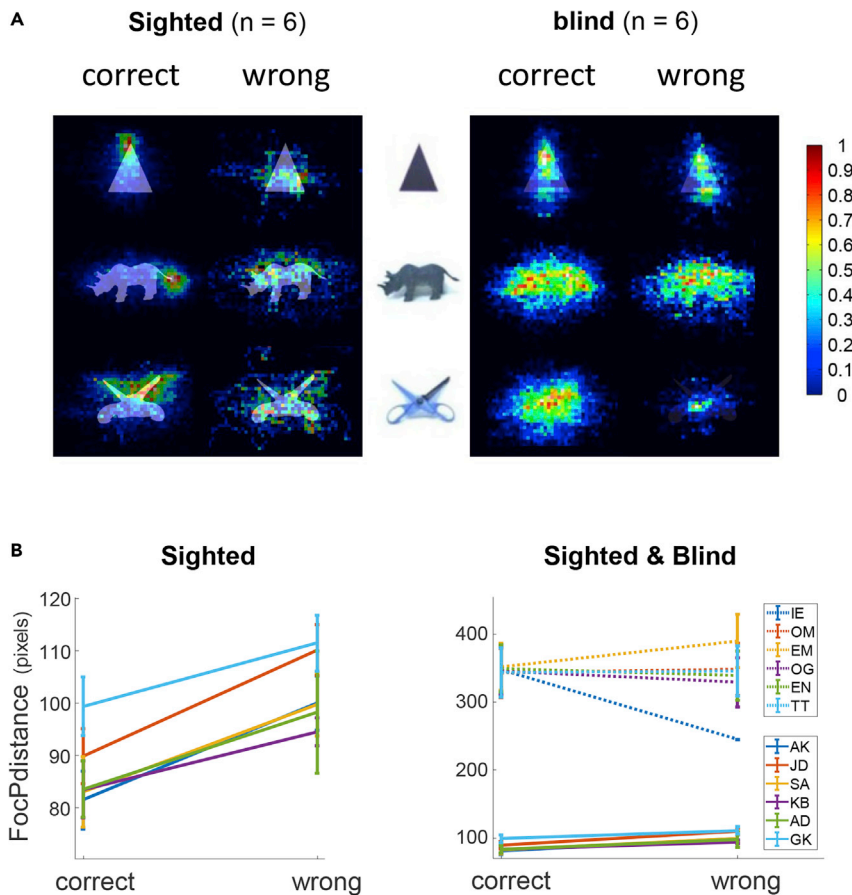


Figure 4. Scanning strategies in sighted and blind participants

(A) Normalized hand visit rates for all correct (left column) and wrong (right column) trials of all sighted (left) and blind participants (right) for one shape and two objects.

(B) Mean distance from the focal point (FocPdistance \pm SEM) in correct and wrong trials for each participant (left, sighted: $p < 0.001$; right, blind: $p = 0.6$; paired t-test).

in a given set required not only convergence on a FocR, if existed, but also identification of the unique spatial information in that region to a level that allows discrimination. We therefore calculated for each object its EP-similarity, which reflects the mean similarity between its EP area and that of each other object in the set (see [Transparent Methods](#)). As expected, the success rates were higher for objects whose EP areas were more spatially unique within their set ([Figure 5E](#); $r = -0.91$, $p = 0.0002$).

Differences in the dynamics of active-sensing between blind and blindfolded sighted participants

We have quantified the differences in the dynamics of active-sensing using three major scanning variables: FocPdistance, FocRtime, VerHozMovement and segmentTime (see [Transparent Methods](#)). For each variable, we calculated the difference between its mean value during the first and second half of the trial, separately for single- and multiple-EP objects ([Figure 6](#)). These comparisons reveal several key differences, the most telling of which are probably those observed when scanning single-EP objects: (i) sighted reduced their FocPdistance in the second half, as a result of convergence—blind did not; (ii) sighted increased their horizontal and vertical scanning movements during the second half—blind did not; (iii) blind shortened their scanning segments during the second half—sighted did not.

Sighted participants exhibited a strong correlation between the amount of time spent on contour following and their mean scanning velocity ([Figure 6B](#), top). Strikingly, no such correlation was observed with our blind participants ([Figure 6B](#), bottom).

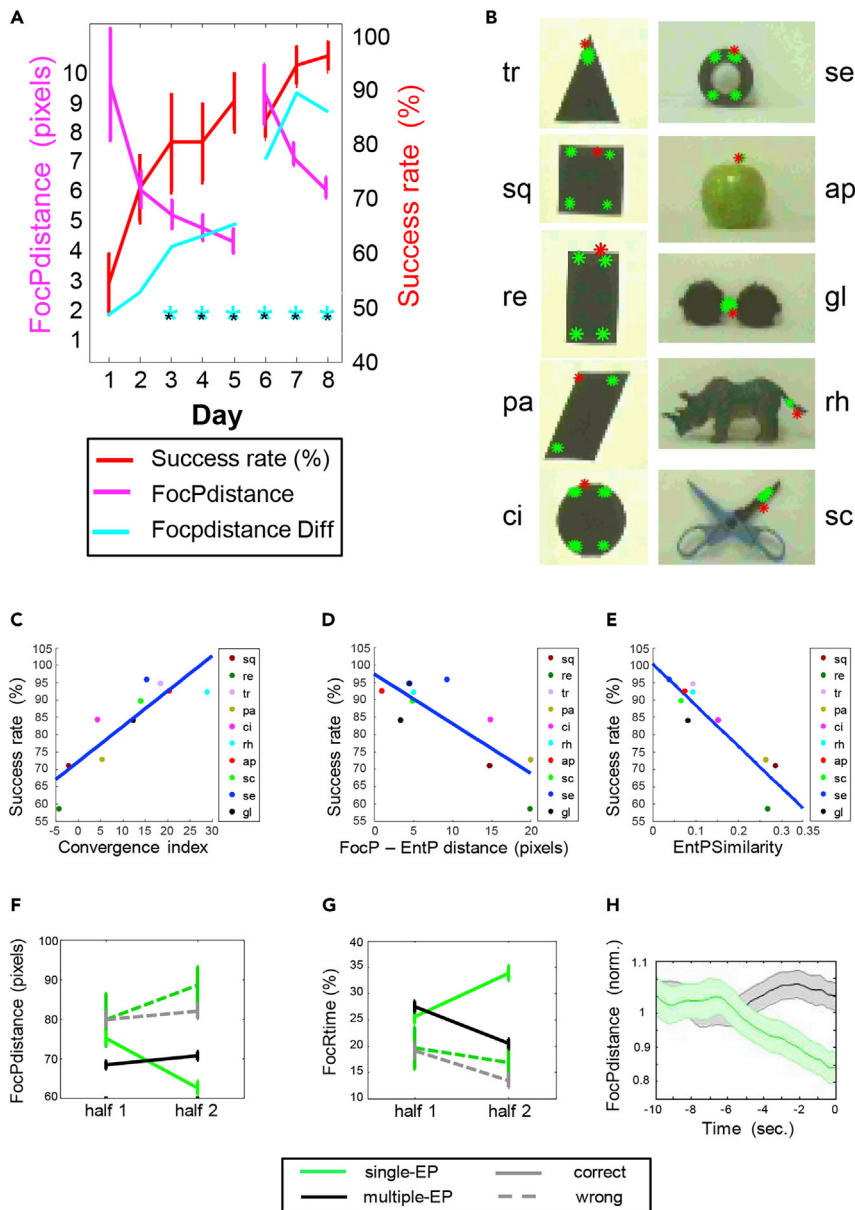


Figure 5. Convergence to focal points (FocPs) in sighted participants

(A) Success rate (red), FocPdistance (magenta) and the difference between the FocPdistance in wrong and correct trials (cyan) as a function of session number (means \pm SEM across all sighted participants, $n = 6$; asterisks denote significant FocPdistance difference [FocPdistance (wrong) – FocPdistance (correct); $p < 0.01$; two-samples t-test].

(B) FocPs (red asterisks) and entropy points (EntPs) (green asterisks) overlaid on objects' snapshots taken from the camera attached to the participants' hands.

(C) Success rate as a function of the convergence index per object; $r = 0.86$, $p = 0.001$.

(D) Overall success rate per object as a function of the distance between FocP and EntP ($r = -0.816$, $p = 0.003$).

(E) Success rate as a function of the EntP's similarity index per object; $r = -0.91$, $p = 0.0002$.

(F–H) Convergence behavior while perceiving objects with single and multiple EntPs. Data averaged separately in the first and second half of each correct and wrong trial, showing (mean \pm SEM): (F) FocPdistance; (G) Percentage of trial time spent within the focal region (FocR). (H) FocPdistance as a function of time preceding the "identification time" ($t = 0$, see [Transparent Methods](#)); curves were normalized to initial values.

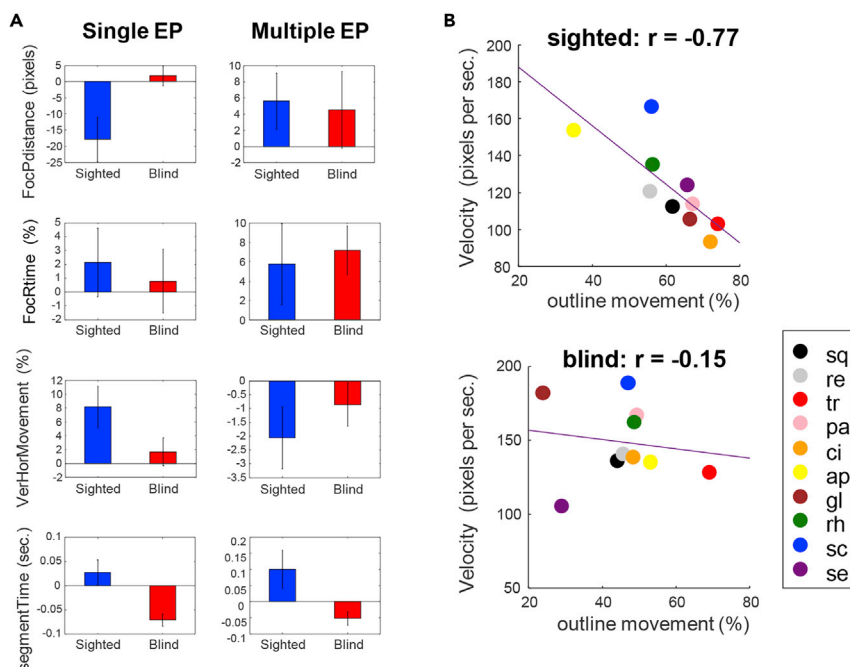


Figure 6. Comparisons of scanning strategies of sighted and blind participants

(A) Convergence in correct trials of single and multiple EP. The differences between the second and the first half of each trial (second minus first) in four variables (variable names appear as ordinate labels; mean \pm SEM across participants, all objects included for each participant).

(B) The relationship between the mean scanning velocity per object and the percentage of time spent on the object's contour (N = 10 objects in each panel).

Discussion

This study provides insights into two domains: sensory substitution and the development of perceptual strategies across modalities for exploring the world and recognizing objects. Applying active-sensing principles to SenSub upgraded performance to an unprecedented level, allowing accurate object recognition within less than 20 s following training of less than 2.5 hr in both blind and blindfolded sighted participants; this performance level retained for at least 20 months in sighted participants. We further show that sighted participants employed a visual-like strategy, focusing on informative sub regions, whereas blind participants typically employed spread scanning—a strategy that is common in tactile perception.

Comparison with other SenSub devices

We compared training time, perception time and accuracy (assessed by the success rate) with 10 previous studies on 2D images and 6 previous studies on 3D objects (Table S1). With 2D images, previous studies showed comparable performances to our ASenSub in either accuracy or perception time (Bach-Y-Rita et al., 1998; Bach-Y-Rita et al., 1969; Tang and Beebe, 1998; van den Doel et al., 2004; Ziat et al., 2005; Chan et al., 2007; Silva et al., 2011). However, no previously reported device comes close to the combined performance achieved with the ASenSub. This is shown by calculating a combined success factor (CBF, see Transparent Methods). The ASenSub's CBF was, on average, 300% and 600% better than previous reports for 2D images and 3D objects, respectively (Table S1). In addition, while all previous devices exhibited poorer accuracy in perceiving 3D objects, they also required much longer training times than our ASenSub (Table S1). Importantly, the unprecedented short perceptual and learning times achieved with the ASenSub may be at a level that can be practical for the blind.

Participants using a device with principles similar to our ASenSub, but in which scanning was constrained to a table, exhibited similar success rates following training, with about twice longer perceptual times (Chan et al., 2007). It is possible that these differences are related to the differences in the motor-sensory context in the two studies—"table-scanning" versus "seeing from a distance". The latter context allows for externalization, which may facilitate learning and perception (Lenay et al., 2003). The different context may also

be a major factor in the choice of motor-sensory strategies. When scanning a table, participants typically choose to primarily focus on contour tracking (Chan et al., 2007). Our sighted participants chose primarily to converge on informative regions of the objects (Figures 4 and 5), a strategy that may better fit remote sensing, such as in natural vision (Land et al., 1999). This feature-oriented strategy was previously reported by participants to be an easy (Auvray et al., 2007) and time saving (Guarniero, 1974; Rovira et al., 2010) strategy.

Studies show better performances (accuracy and recognition time) in haptic task for visually impaired participants than for sighted controls (Picard et al., 2014). Our research shows similar success levels (Figure 2B) but different scanning patterns for the sighted and the blind participants (Figures 3B, 4, and 6). Thus, for example, sighted participants typically converged on visually informative focal points whereas congenital blinds preferred more uniform scanning. This comparison suggests that active sensory strategies are developed according to available resources and appear to prefer functionality over modality-specific cues (Pascual-Leone and Hamilton, 2001).

Active strategies

Information is largely affected by the scanning pattern (Findlay and Brown, 2006; Noton and Stark, 1971; Yang et al., 2016). Strategy differences were previously shown between blind and sighted participants while recognizing an object by touch. Some of these studies claimed for the superiority, via adaptive strategies, of the blind over the sighted (D'anggiulli and Kennedy, 2001; Davidson, 1972; Davidson and Whitson, 1974; Russier, 1999; Rovira et al., 2011); in general, sighted strategies were shown to contain more focused and sequential scanning than those of the blind.

Consistent with the above, the strategy our sighted participants developed was based on denser scanning of informative regions of the objects, resembling natural vision (Reinagel and Zador, 1999; Krieger et al., 2000; Parkhurst and Niebur, 2003; Mack et al., 2003; O'Bryan and Boersma, 1971). Consequently, objects in which the spatial pattern of information had a single distinct informative region were easier to perceive. Sequences of palpation of our sighted participants, when not dwelling in a focal region (FocR; Figure 4), included scanning along and across the object's contour (Figure 3A), in a manner resembling ocular trajectories in human participants with artificially induced tunnel vision (Gruber and Ahissar, 2020). These latter movements probably provided valuable information about the object, information that was likely verified during the focal scanning of FocRs.

The general transition from one informative region to another (e.g., contours and FocRs) may resemble typical trajectories of eye saccades (Walker-Smith et al., 1977; Ko et al., 2010), where the selection of targets seems to be affected by bottom-up feature saliency and top-down heuristics (Navalpakkam and Itti, 2007), as was often evident from our participants' introspective reports (Video S1). In most cases, when palpating objects with a single entropy peak (EP; Figure 5B), our participants were able to find the regions with the highest entropy (largest amount of spatial information) even though our ASenSub did not contain any peripheral information.

The acquisition of local sensory information likely did not involve any conscious planning, as our participants' introspective reports did not contain any description of how they identified local features, which suggests that local features were perceived via low-level motor-sensory perceptual routines. Yet, our data cannot tell whether the improvement allowed by our ASenSub approach was due to a better focus on relevant sensory features, or rather on the better adequacy of the protocol with the closed-loop nature of somatosensation.

Comparing blind and sighted strategies

Understanding of object material is dominant in touch while understanding geometric characteristic is dominant in vision (Lederman and Klatzky, 1990). Therefore, visual imagination helps in geometric characteristic task but not in material tasks (Heller, 1989). The most efficient way to get object characteristic by touch is grasping and lifting, which provide information on the object's shape and material (Lederman and Klatzky, 1990). Specifically, it was shown that simultaneous sensing of several contour points is preferable over contour following with one finger (Abravanel, 1968; Hatwell, 1960). These findings are in agreement with the scanning strategy developed by our blind participants trying to encompass all sides of the object rather than focusing on one region of the object (Figure 4).

Blind probably achieve superior touch sensitivity via learning and cross-modal plasticity (Grant et al., 2000; Wong et al., 2011). This superiority is expressed in accuracy levels, learning speed, or both (Goldreich and Kanics, 2003; Norman and Bartholomew, 2011; Withagen et al., 2012; Rovira et al., 2011; Röder et al., 1996). As both our sighted and blind participants approached perfect accuracy, the only meaningful comparisons here are learning speed and perceptual speed. Our blind participants learned as fast as our sighted participants (Figures 2B and 2C). As for perceptual speed, our results reveal larger variability within our blind participants (Figures 2B and 2C, EXP1), with some of them exhibiting extremely fast perception. Possibly, these differences are related to the different exploratory strategies used by our sighted and blind participants, raising the possibility that, for the sighted participants, the focal strategy was more reliable and repeatable than global scanning when attempting to recognize familiar objects from familiar view angles. In general, we think that the differences revealed here might be explained by the different spatial perception in relation to the moving perceiving body (Morris, 2004).

One methodological issue should be discussed in this context. Before the experiments began, the sighted participants saw the objects (EXP 1–3), while blind participants tactually palpated a one-to-one physical model of the 2D shapes (in EXP 1) or the 3D objects themselves (in EXP 2). This procedural difference was taken in order to allow the participants to first recognize the objects in the manner that they were used to, and on which their natural style of report was based on. Yet, a potential uncontrolled bias might be introduced by this difference. Therefore, in EXP 4 the same sighted participants were tested with novel shapes that they did not see before, and they were only allowed to tactually palpate their one-to-one physical models before the experiment began (the same procedure that was used with the blind participants in EXP 1&2).

Conclusive summary

In summary, we show here that sensory substitution that allows the employment of natural motor-sensory perceptual strategies (O'Regan and Noe, 2001; Ahissar and Arieli, 2001) can shorten acquisition time and training time enormously. Natural motor-sensory strategies differ between blind and sighted subjects. Yet, both groups could rapidly adapt the device to their natural strategies, suggesting that the perception of objects is mediated by experience-dependent strategies of sensor motion. Moreover, the same participants could perceive the same objects via different motor-sensory patterns (in different trials), suggesting that their perception made use of adaptive closed loop mechanisms (Uexkull, 1926; Wiener, 1949; Powers, 1973; Ahissar and Vaadia, 1990; Saig et al., 2012). The rapid learning of cross-modal perception observed here, and specifically the indications for a rapid adaptation of visually based strategies for tactile sensation, suggest that the principles of visual and tactile perception share critical components. The striking similarity between the learning rates, accuracies and perceptual speeds of congenitally blind and blindfolded sighted subjects means that prior vision is not a pre-requisite here. The fast training and perceptual speeds enabled by these strategies suggest that active-sensing based device may be instrumental in rendering sensory substitution practical for daily use by blind people. To be practical, the large prototype that we used in the laboratory (for obtaining adequate stimulus control and motion tracking) should be replaced by a miniature ASenSub that either is added to, or will be part of a walking cane or a glove-like wearable item. Such technology will enable easier involvement of blind people in our society.

Limitations of the study

The most important limitation of this study is probably the lack of testing in natural, daily life conditions. Such testing requires the implementation of the ASenSub using wearable technology, a technology that is not yet available for tactile actuation at a reasonable resolution. Another important limitation is the lack of reference to peripheral sensing. Inferring from vision, peripheral sensation is expected to be crucial for an efficient use of ASenSub in natural conditions. The implementation of peripheral sensation in ASenSub is challenging, and requires a dedicated project that will converge on efficient combinations of receptor architecture, activation principles and fovea-periphery interplay. In addition, the scope of the current study is limited in the spectrums of object types, sizes and distances tested.

Resource availability

Lead contact

Further information and requests for resources should be directed to the lead contact, Amos Arieli (amos.arieli@weizmann.ac.il).

Materials availability

This study did not generate new unique reagents/materials.

Data and code availability

The dataset and code supporting the current study are available from the lead contact upon reasonable request.

Methods

All methods can be found in the accompanying [Transparent methods supplemental file](#).

Supplemental information

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101918>.

Acknowledgments

We thank Avi Saig for helping in setting up the system and advices, Yonatan Kra for helpful discussions and advices, Daniel Goldian for great technical support, Shira Ozana for advices, and Roy Harpaz for statistical consultation. This research received funding from the European Research Council (ERC) under the EU Horizon 2020 Research and Innovation Program (grant agreement No 786949). E.A. holds the Helen Diller Family Professorial Chair of Neurobiology. A.A. holds the Sam and Frances Belzberg Chair in Memory and Learning.

Author contributions

Conceptualization, A.A. and E.A. and Z.K.Y.; **Experimental Design**, A.A. and E.A. and Y.Z.K.; **Methodology**, A.A. and E.A. and S.G. and Y.Z.K.; **Data Collection**, Y.Z.K. and S.G.; **Software/Analysis**, A.A. and Y.Z.K.; **Writing**, A.A. and E.A. and Y.Z.K.; **Funding Acquisition**, A.A. and E.A.; **Supervision**, A.A. and E.A.

Declaration of interests

The authors declare no competing interests.

Received: August 17, 2020

Revised: October 25, 2020

Accepted: December 7, 2020

Published: January 22, 2021

References

- Abравanel, E. (1968). Intersensory integration of spatial position during early childhood. *Percept. Mot. skills* 26, 251–256.
- Ahissar, E., and Arieli, A. (2001). Figuring space by time. *Neuron* 32, 185–201.
- Ahissar, E., and Arieli, A. (2012). Seeing via miniature eye movements: a dynamic hypothesis for vision. *Front. Comput. Neurosci.* 6, 89.
- Ahissar, E., and Assa, E. (2016). Perception as a closed-loop convergence process. *eLife* 5, e12830.
- Ahissar, E., and Vaadia, E. (1990). Oscillatory activity of single units in a somatosensory cortex of an awake monkey and their possible role in texture analysis. *Proc. Natl. Acad. Sci. U S A* 87, 8935–8939.
- Amedi, A., Stern, W.M., Camprodon, J.A., Bermpohl, F., Merabet, L., Rotman, S., Hemond, C., Meijer, P., and Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat. Neurosci.* 10, 687–689.
- Arno, P., Capelle, C., Wanet-Defalque, M.-C., Catalan-Ahumada, M., and Veraart, C. (1999). Auditory coding of visual patterns for the blind. *Perception* 28, 1013–1030.
- 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, 416–430.
- Bach-Y-Rita, P., Collins, C.C., Saunders, F.A., White, B., and Scadden, L. (1969). Vision substitution by tactile image projection. *Nature* 221, 963–964.
- Bach-Y-Rita, P., Kaczmarek, K.A., Tyler, M.E., and Garcia-Lara, J. (1998). Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note. *J. Rehabil. Res. Dev.* 35, 427–430.
- Borji, A., Sihite, D.N., and Itti, L. (2013). What stands out in a scene? A study of human explicit saliency judgment. *Vis. Res.* 91, 62–77.
- Brown, D., Macpherson, T., and Ward, J. (2011). Seeing with sound? exploring different characteristics of a visual-to-auditory sensory substitution device. *Perception* 40, 1120–1135.
- Cascio, C., and Sathian, K. (2001). Temporal cues contribute to tactile perception of roughness. *J. Neurosci.* 21, 5289–5296.
- Chan, J.S., Maucher, T., Schemmel, J., Kilroy, D., Newell, F.N., and Meier, K. (2007). The virtual haptic display: a device for exploring 2-D virtual shapes in the tactile modality. *Behav. Res. Methods* 39, 802–810.
- Collins, C.C., and Bach-Y-Rita, P. (1973). Transmission of pictorial information through the skin. *Adv. Biol. Med. Phys.* 14, 15.
- Connor, C.E., and Johnson, K.O. (1992). Neural coding of tactile texture: comparison of spatial

and temporal mechanisms for roughness perception. *J. Neurosci.* 12, 3414–3426.

D'anggiulli, A., and Kennedy, J.M. (2001). Children's tactual exploration and copying without vision. *Int. J. Rehabil. Res.* 24, 233–234.

Davidson, P.W. (1972). Haptic judgments of curvature by blind and sighted humans. *J. Exp. Psychol.* 93, 43.

Davidson, P.W., and Whitson, T.T. (1974). Haptic equivalence matching of curvature by blind and sighted humans. *J. Exp. Psychol.* 102, 687.

Doel, K.V.D., Smilek, D., Bodnar, A., Chita, C., Corbett, R., Nekrasovski, D., and Mcgreneire, J. (2004). Geometric shape detection with soundview. In Proceedings of the 10th International Conference on Auditory Display (ICAD2004).

Findlay, J.M., and Brown, V. (2006). Eye scanning of multi-element displays: II. Saccade planning. *Vis. Res.* 46, 216–227.

Gamzu, E., and Ahissar, E. (2001). Importance of temporal cues for tactile spatial- frequency discrimination. *J. Neurosci.* 21, 7416–7427.

Goldreich, D., and Kanics, I.M. (2003). Tactile acuity is enhanced in blindness. *J. Neurosci.* 23, 3439–3445.

Grant, A.C., Thiagarajah, M.C., and Sathian, K. (2000). Tactile perception in blind Braille readers: a psychophysical study of acuity and hyperacuity using gratings and dot patterns. *Percept. Psychophys.* 62, 301–312.

Gruber, L.Z., and Ahissar, E. (2020). Closed loop motor-sensory dynamics in human vision. *PLoS One* 15, e0240660.

Guarniero, G. (1974). Experience of tactile vision. *Perception* 3, 101–104.

Hatwell, Y. (1960). Étude de quelques illusions géométriques tactiles chez les aveugles. *L'année psychologique* 60, 11–27.

Heller, M.A. (1989). Picture and pattern perception in the sighted and the blind: the advantage of the late blind. *Perception* 18, 379–389.

Hollins, M., and Bensmaïa, S.J. (2007). The coding of roughness. *Can. J. Exp. Psychol.* 61, 184.

Kelso, J.S. (1997). *Dynamic Patterns: The Self-Organization of Brain and Behavior* (MIT press).

Ko, H.K., Poletti, M., and Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. *Nat. Neurosci.* 13, 1549–1553.

Krieger, G., Rentschler, I., Hauske, G., Schill, K., and Zetsche, C. (2000). Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. *Spat. Vis.* 13, 201–214.

Land, M.F., Mennie, N., and Rusted, J. (1999). Eye movements and the roles of vision in activities of daily living: making a cup of tea. *Perception* 28, 1311–1328.

Lederman, S.J., and Klatzky, R.L. (1990). Haptic classification of common objects: Knowledge-driven exploration. *Cogn. Psychol.* 22, 421–459.

Lenay, C., Gapenne, O., Hanneton, S., Marque, C., and Geouelle, C. (2003). Sensory substitution: Limits and Perspectives. In *Touching for Knowing, Cognitive Psychology of Haptic Manual Perception* (John Benjamins Publishing Company).

Mack, M., Castelhamo, M.S., Henderson, J.M. & Oliva, A. What the visual system sees: the relationship between fixation positions and image properties during a search task in real-world scenes. *Proc. Ann. Object, Perception, Attention, and Memory Conf.* 2003.

Mackevicius, E.L., Best, M.D., Saal, H.P., and Bensmaïa, S.J. (2012). Millisecond precision spike timing shapes tactile perception. *J. Neurosci.* 32, 15309–15317.

Maidenbaum, S., Abboud, S., and Amedi, A. (2013). Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation. *Neurosci. Biobehav. Rev.* 41, 3–15.

Mannan, S.K., Kennard, C., and Husain, M. (2009). The role of visual salience in directing eye movements in visual object agnosia. *Curr. Biol.* 19, R247–R248.

Morris, D. (2004). *The Sense of Space* (SUNY Press).

Navalpakkam, V., and Itti, L. (2007). Search goal tunes visual features optimally. *Neuron* 53, 605–617.

Norman, J.F., and Bartholomew, A.N. (2011). Blindness enhances tactile acuity and haptic 3-D shape discrimination. *Atten. Percept. Psychophys.* 73, 2323–2331.

Noton, D., and Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science* 171, 308–311.

Nye, P.W., and Bliss, J.C. (1970). Sensory aids for the blind: a challenging problem with lessons for the future. *Proc. IEEE* 58, 1878–1898.

O'Bryan, K.G., and Boersma, F.J. (1971). Eye movements, perceptual activity, and conservation development. *J. Exp. Child Psychol.* 12, 157–169.

O'Regan, J.K., and Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–973, discussion 973–1031.

Parkhurst, D.J., and Niebur, E. (2003). Scene content selected by active vision. *Spat. Vis.* 16, 125–154.

Pascual-Leone, A., and Hamilton, R. (2001). The metamodal organization of the brain. *Prog. Brain Res.* 134, 427–446.

Pei, Y.-C., Hsiao, S.S., Craig, J.C., and Bensmaïa, S.J. (2010). Shape invariant coding of motion direction in somatosensory cortex. *PLoS Biol.* 8, e1000305.

Picard, D., Albaret, J.-M., and Mazella, A. (2014). Haptic identification of raised-line drawings when

categorical information is given: a comparison between visually impaired and sighted children. *Psicologica: Int. J. Methodol. Exp. Psychol.* 35, 277–290.

Powers, W.T. (1973). Feedback: beyond behaviorism. *Science* 179, 351–356.

Ptito, M., Moesgaard, S.M., Gjedde, A., and Kupers, R. (2005). Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind. *Brain* 128, 606–614.

Reinagel, P., and Zador, A.M. (1999). Natural scene statistics at the centre of gaze. *Netw. Comput. Neural Syst.* 10, 341–350.

Röder, B., Rösler, F., Hennighausen, E., and Näcker, F. (1996). Event-related potentials during auditory and somatosensory discrimination in sighted and blind human subjects. *Cogn. Brain Res.* 4, 77–93.

Rovira, K., Deschamps, L., and Baena-Gomez, D. (2011). Mental rotation in blind and sighted adolescents: the effects of haptic strategies. *Eur. Rev. Appl. Psychol.* 61, 153–160.

Rovira, K., Gapenne, O., and Ammar, A.A. (2010). Learning to recognize shapes with a sensory substitution system: a longitudinal study with 4 non-sighted adolescents. In *2010 IEEE 9th International Conference on Development and Learning (IEEE)*, pp. 1–6.

Russier, S. (1999). Haptic discrimination of two-dimensional raised-line shapes by blind and sighted adults. *J. Vis. Impairment Blindness* 93, 421–426.

Saig, A., Gordon, G., Assa, E., Arieli, A., and Ahissar, E. (2012). Motor-sensory confluence in tactile perception. *J. Neurosci.* 32, 14022–14032.

Sathian, K. (1989). Tactile sensing of surface features. *Trends Neurosci.* 12, 513–519.

Schumann, F., Einhäuser, W., Vockeroth, J., Bartl, K., Schneider, E., and König, P. (2008). Salient features in gaze-aligned recordings of human visual input during free exploration of natural environments. *J. Vis.* 8, 12.

Silva, P.M., Pappas, T.N., Atkins, J., and West, J.E. (2011). Perceiving graphical and pictorial information via touch and hearing. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE)*, pp. 2292–2295.

Striem-Amit, E., Dakwar, O., Reich, L., and Amedi, A. (2012). The large-scale organization of "visual" streams emerges without visual experience. *Cereb. Cortex* 22, 1698–1709.

Tang, H., and Beebe, D.J. (1998). A microfabricated electrostatic haptic display for persons with visual impairments. *IEEE Trans. Rehabil. Eng.* 6, 241–248.

Uexkull, J.V. (1926). *Theoretical Biology* (London, K. Paul, Trench, Trubner & co. Ltd).

van den Doel, K., Smilek, D., Bodnar, A., Chita, C., Corbett, R., Nekrasovski, D., and Mcgreneire, J. (2004). Geometric Shape Detection with Soundview (Georgia Institute of Technology).

Visell, Y. (2009). Tactile sensory substitution: models for enaction in HCI. *Interact Comput.* 21, 38–53.

Walker-Smith, G.J., Gale, A.G., and Findlay, J.M. (1977). Eye movement strategies involved in face perception. *Perception* 6, 313–326.

Weber, A.I., Saal, H.P., Lieber, J.D., Cheng, J.-W., Manfredi, L.R., Dammann, J.F., and Bensmaia, S.J. (2013). Spatial and temporal codes mediate the tactile perception of natural textures. *Proc. Natl. Acad. Sci. U S A* 110, 17107–17112.

Wiener, N. (1949). *Cybernetics* (John Wiley & Sons).

Withagen, A., Kappers, A.M., Vervloed, M.P., Knoors, H., and Verhoeven, L. (2012). Haptic object matching by blind and sighted adults and children. *Acta Psychol.* 139, 261–271.

Wong, M., Gnanakumaran, V., and Goldreich, D. (2011). Tactile spatial acuity enhancement in blindness: evidence for experience-dependent mechanisms. *J. Neurosci.* 31, 7028–7037.

Yang, S.C.-H., Lengyel, M., and Wolpert, D.M. (2016). Active sensing in the categorization of visual patterns. *Elife* 5, e12215.

Yoshioka, T., Craig, J.C., Beck, G.C., and Hsiao, S.S. (2011). Perceptual constancy of texture roughness in the tactile system. *J. Neurosci.* 31, 17603–17611.

Ziat, M., Gapenne, O., Stewart, J. & Lenay, C. A Comparison of Two Methods of Scaling on Form Perception via a Haptic Interface. 7th international conference on Multimodal interfaces, October 2005, Toronto, Italy. 236-243.

Zilbershtain-Kra, Y., Arieli, A., and Ahissar, E. (2015). Tactile substitution for vision. *Scholarpedia* 10, 32457.

iScience, Volume 24

Supplemental Information

**Active sensory substitution allows
fast learning via effective
motor-sensory strategies**

Yael Zilbershtain-Kra, Shmuel Graffi, Ehud Ahissar, and Amos Arieli

Supplemental Table and legend

Reference	SenSub vision to:	N	Chance level	Success Begin %	Success End %	Success avg. %	IdenTim Begin (sec.)	IdenTim End (sec.)	Time limit (sec.)	Training time (hrs.)	(CBF)
2D											
Bach –Y- Rita, 1969	touch	6	33	—	—	—	—	—	—	1	—
Tang, 1998	touch	4	33	61.8* ¹	76.4* ¹	72.1	—	—	—	0.5+?	—
Doel, 2004	auditory	60	16.66	66.2	—	66.2	—	—	90	0.33	0.55
Doel, 2004	auditory	12	25	75	—	75	47	—	90	0.75	0.55
Ziat, 2005	touch	10	33	—	—	77	—	—	90	—	0.48
Chan, 2007	touch	10	25	46	84	65	60	60	60+	1.5	0.66
Brown, 2011	auditory	18	25	28	34	31	—	—	—	3	—
Silva, 2011	auditory	10	33	—	—	83.3	—	—	—	—	—
Pietrzak, 2009	touch	8	10	—	—	90	—	95.95	180	—	0.44
Bach –Y- Rita, 1998	touch	5	33	—	—	79.8	—	—	—	—	—
<i>Sighted EXP 1</i>	<i>touch</i>	<i>6</i>	<i>20</i>	<i>53</i>	<i>87</i>	<i>74</i>	<i>27</i>	<i>17</i>	<i>26</i>	<i>2.5</i>	<i>2.07</i>
<i>Sighted EXP 4</i>	<i>touch</i>	<i>4</i>	<i>20</i>	<i>100</i>	<i>100</i>	<i>99.2</i>	<i>18.6</i>	<i>10.9</i>	<i>30</i>	<i>4*² + 3</i>	<i>2.64</i>
<i>Blind EXP 1</i>	<i>touch</i>	<i>6</i>	<i>20</i>	<i>40</i>	<i>74.16</i>	<i>62.7</i>	<i>27.48</i>	<i>25.61</i>	<i>30</i>	<i>4</i>	<i>1.42</i>
<i>Blind EXP 2</i>	<i>touch</i>	<i>6</i>	<i>20</i>	<i>80</i>	<i>90.83</i>	<i>84.44</i>	<i>23.7</i>	<i>20.67</i>	<i>30</i>	<i>5.5</i>	<i>2.14</i>
3D											
Bach –Y- Rita, 1969	touch	6	4	—	—	—	—	5-20	—	11	—
Shinohara, 1998	touch	6	10	—	—	—	—	—	—	—	—
Auvray, 2007* ³	auditory	6	10	—	—	* ⁴	—	—	120	11	—
Auvray, 2007* ⁵	auditory	6	10	—	—	* ⁶	—	—	120	15	—
Brown, 2011	auditory	18	25	60	60	60	90	49	90	4	0.39
Brown, 2011	auditory	24	11	44	44	44	63	43	90	6	0.36
<i>Sighted EXP 2</i>	<i>touch</i>	<i>6</i>	<i>20</i>	<i>83</i>	<i>95</i>	<i>90.56</i>	<i>25</i>	<i>22</i>	<i>30</i>	<i>4</i>	<i>2.35</i>

Table S1. Our participants performed remarkably better than what was previously reported in 2D shapes and 3D objects. Related to Discussion: ‘Comparison with other SenSub devices’. The CBF (right column, see STAR Methods) was calculated for all studies which reported about the success rate, the chance level and the time limit. **Table headers:** N, number of participants. Chance level, relate to the number of different objects which were presented for identification. Success Begin, averaged success rate in the first experimental session. Success End, averaged success rate in the last experimental session. Success avg., averaged success rate in all experiments. IdenTim Begin, averaged identification time (sec.) in the first experimental session. IdenTim End, averaged identification time (sec.) in the last experimental session. Time limit, for identification (sec.). In ***Bold-Italic***, this study’s EXP 1 to 4. **Comments:** In all references we present the best result except in *¹ where only the averaged result were published. *², these participants were trained for 4 hours ~2 years before. *³, Exp. 3 in Auvray, 2007. *⁴, success rate was not reported in this study, however ~1.6 objects on average were reported before the correct answer and 12% failed to make a response before the trial was terminated. *⁵, Exp. 4 in Auvray, 2007. *⁶, ~1.7 objects on average were reported before the correct answer and 13% failed to make a response before the trial was terminated.

Transparent Methods

Participants

Six healthy sighted participants aged 23-33 (mean 26.6 ± 4) of both genders (4 males and 2 females) and six healthy congenitally blind participants aged 28-40 (mean 32.5 ± 4) of both genders (1 males and 5 females) volunteered to participate in the experiment. All sighted participants had normal or corrected-to-normal vision and all but one were right handed. All blind participants were congenitally blind (blindness onset 0-2 years old) and were all brail readers. Tasks were carried out with participant's understanding and consent. Informed consent forms were obtained from all the participants, in accordance with the approved Declaration of Helsinki for this project. The Participants were paid for their participation.

Apparatus

The Active-sensing based sensory substitution device (ASenSub), which was attached to the participant's hand, included a sensor and an actuator (**Fig. 1**). The sensor was a miniature video camera (#2 “active camera”, WAT-230A, Watec, Japan; 30 frames per second, fps). The actuator (#1 VTMouse; Tactile World, Israel), had three touch arrays for three fingers, each with 32 pins (4×8); the height of each pin could be set to one of four different levels. The set of 12×8 pins was mapped to a set of 192×128 pixels at the central field of the active camera (**Fig. 1** bottom right), such that each pin was activated by the mean brightness of an array of 16×16 video pixels. This central region covered $2.6^\circ \times 1.7^\circ$ of the visual field, corresponding to an area of 4.5×3.0 cm at a distance of 1 m from the ASenSub (where the objects were presented, **Fig. 1** #3). The setup also included a fixed camera (#5 DCC1645C, USB camera, Thorlabs, Germany), which was installed immediately below the location of the ASenSub and captured one full-field frame for each object before the experiment begun.

Active-sensing operational principles were enforced by two major design principles. First, the same body organ that moved to scan the environment (in our experiment - the hand) received the tactile stimulation. Second, no artificial active components (such as mechanical vibrations) were added; the elevation level of each actuator pin was directly

related to the level of illumination captured by its corresponding pixel at the camera, without adding any vibration (Fig. 1 bottom right).

The ASenSub was programmed and operated using MATLAB software. The weight of the ASenSub was balanced in part by a spring balancer (VGL, SB02, capacity range 1.0-2.0 kg) that was attached to an X-Y track systems on the ceiling (coverage distance of 1 meter), enabling effortless 3D hand movement (**Fig. 1 #4**).

Objects and procedure

Sighted participants were first invited to eight experimental sessions (EXP 1&2) followed by 6 experimental sessions ~20 months later (EXP 3&4). Blind participants were invited to 8 experimental sessions (EXP 1&2).

In all experiments, the participants were asked to identify one object out of 5 possible objects. In EXP 2 the objects were 3D physical objects (e.g., an apple in **Fig. 1 #3**) while in all the other experiments the objects were 2D black shapes that were presented on a computer screen; throughout this article the generic term “object” refers to both types. The background of the objects was always white (white map on a table near a white wall and a white screen background, respectively). The distance between the object and the ASenSub was around 1 m. A streamer was stretched in front of the participants to prevent a shortening of the distance.

In EXP 1-3 the sighted participants saw the objects before the experiment began. In EXP 4 the participants never saw the objects. Instead, they could palpate a one-to-one physical model of the 2D shapes before the experiment began. Blind participants were also allowed to palpate those objects before each experiment began. Each participant was given a general explanation about the ASenSub system at the beginning of the first session, and was able to practice the system for up to 10 minutes at the beginning of each experiment with novel objects.

In each session, they were tested in 20 trials with a short rest break of 5 minutes between the first and the last ten trials. During the break, participants were allowed to take off the ASenSub and their blindfold and to sit on a chair without seeing the objects. In each trial

participants were presented with one randomly-chosen object and asked to identify it as fast as possible. During the trials they were encouraged to voice their thoughts and strategies; their reports were recorded by a small microphone that was attached to their chest (**Fig. 1 #6**). The trial ended with the participant's declaration of identification, or after the time limit was passed (whichever came first). At the end of each trial, the participants reported their identification and their confidence level (0-100 percent). A detailed feedback was then given (whether their answer was correct or wrong and what the correct answer was).

EXP 1. The first experiment - 5 sessions. Set of objects: 5 2D images (square, rectangle, triangle, parallelogram and circle; **Fig. 2A**). Trial time limit was 30 seconds during the first 3 sessions and 20 seconds during the last 2 sessions.

EXP 2. Commenced immediately after EXP 1 - 3 sessions. Set of objects: 5 3D physical objects (glasses, sellotape, apple, scissors and a rhinoceros toy; **Fig. 2A**). Trial time limit was 30 seconds.

EXP 3. Commenced for the sighted participants 20 - 20.5 months after EXP 2 - 3 sessions. Set of objects: as in EXP 1. Trial time limit was 30 seconds.

EXP 4. Commenced for the sighted participants immediately after EXP 3 - 3 sessions. Set of objects: five 2D images (heart, trapezoid, moon, arrow and the digit 1; **Fig. 2A**). Trial time limit was 30 seconds.

Each pin in the ASenSub covered an area of 0.37×0.37 cm at a distance of 1 m from the ASenSub (Fig. 1). 2D images spanned 6.7-10.0 cm horizontally (mean = 8.3 cm; 4.7°) and 8.6-12.1 cm vertically (mean = 10.3 cm; 5.9°) whereas 3D objects spanned 8.3-17.9 cm horizontally (mean = 14.2 cm; 8.1°) and 6.7-13.1 cm vertically (mean = 9.5 cm; 5.5°).

Data analysis

Data acquired from the active camera was analyzed in order to characterize the motor-sensory strategies applied by our participants.

Scanning trajectory. A MATLAB code was developed for tracking the scanning trajectory across the object. The algorithm was based on computing the spatial correlation between

the frame taken from the fixed camera and every 2nd frame taken from the active camera. The two frames were cropped around the object, converted to binary B&W images and conditioned (using MATLAB functions ‘imfill’, ‘bwareaopen’, ‘imresize’ and ‘imdilate’) before correlation. Frames taken from the active camera (“dynamic frames”) with < 10 pixels sampling the object were skipped, and the trajectory was interpolated over them. The code was executed in a semi-automatic manner; the resulting trajectory was examined while constructed and conditioning parameters were adapted to address significant changes such as substantial changes in viewing distance.

Categorization of scanning movements. Scanning trajectories of each trial were parsed into movement sequences (SEQs), in relation to the contour of the object’s cross section and the hand’s “gaze” (defined as the center of the dynamic frame). First, the video of the dynamic camera was down-sampled to 15 fps by taking every 2nd frame. Then four SEQ types – “contour”, “horizontal”, “vertical” and “other” (**Fig. 3**) – were defined using the following heuristic algorithms. Contour SEQs included sequences of at least 15 frames (1 s) whose gazes distanced < D1 pixels from the nearest contour of the object. D1 was set to 20 pixels for 2D images and 15 pixels for 3D physical objects (the latter occupied less pixels than the formers, with a mean ratio of 3:4 between the long axes of their contours). Adjacent contour-SEQs were combined if the distance between the last frame in the leading SEQ and the first frame in the following SEQ was < D2 (D2 = 25 pixels for 2D images and 20 pixels for 3D physical objects). *VerHozMovement* was calculated for each trial as the percent of the number of frames that were horizontal and vertical SEQ out of the total number of frames in that period. Horizontal and vertical SEQs included sequences of at least 10 frames that were not defined within any contour SEQ and whose “movement slope” was consistently larger (“vertical”) or smaller (“horizontal”) than 1. The movement slope was computed as the absolute shift (from the previous frame) of the gaze in the vertical direction divided by that along the horizontal direction. Five-point boxcar averaging was applied to horizontal and vertical SEQs.

Visit maps, focal points (FocPs) and focal regions (FocRs). Scanning trajectories were averaged and normalized per object for all trials of each participant and each session separately, or for all participants and all sessions together, in bins of 10×10 pixels. Each bin was assigned a ‘visit rate’ (VR), which was the total number of visits in that bin divided

by trial duration (T). Each visit map was normalized by dividing each of its values by the maximal VR value in that map.

The region with the highest rate values in each map was quantified as the region in which all bins had visit rates > 0.95 ; regions containing one such bin surrounded by bins with rates < 0.55 were excluded. The center of mass of this region was defined as the primary FocP of that map. A secondary FocP was defined using the same procedure after removing the primary FocP from the map. The FocR was defined as the region of pixels included within a radius of 40 pixels from the FocP.

Entropy map and entropy point (EP). The EP was defined as the pixel in the object's cross section around which the entropy of the tactile image is maximal. The entropy of the tactile image was computed for each object from a representative dynamic frame as follows. A binary image was constructed, where the pixels having illumination levels inducing maximal pin deflection (see "Active-sensing based sensory substitution" above) were set to "1" and all other pixels set to "0". The image was conditioned (using MATLAB function 'imfill') to reduce the effect of noisy sampling. The local entropy of each pixel was computed using the MATLAB function 'entropyfilt', for a neighborhood of 3*3 pixels.

Identification moment. Each trial was ended by the experimenter pressing a button immediately after hearing the participant reporting the identification (verbally). An upper bound for the trial-time in which the participant identified the object (t_{id}) was assessed as $t_{id} = t_R - t_E$, where t_R was the time of button-press by the experimenter, and t_E is the estimated reaction time of the experimenter. t_E was assessed as 0.55 ± 0.04 s (Mean \pm SE; $n = 20$ independent measurements).

Additional analysis variables:

EPsimilarity. The similarity of the EP between two objects (EPsimilarity) was determined by calculating the mutual information between the 'tactile' areas (8*12 pixels) around the EP. For multiple-EP objects, the EP used for this calculation was the one closest to the FocP. For each object, the similarity index was the average of its pair-wise similarities with all other set members.

FocPdistance. The FocPdistance of a pixel in a given object is the Euclidian distance between the pixel and the closer FocP in that object (out of the primary and secondary, if exists, FocPs).

FocP-EP distance. This is the Euclidian distance between the FocP (translated to the entropy map using a correlation procedure as in “Visit maps, focal points (FocPs) and focal regions (FocRs)” above) and its nearest EP.

FocRtime. The percent time spent in scanning an FocR (FocRtime) was calculated for each half a trial as the percent of the number of frames that were part of contour SEQs and were within the FocR during that period out of the total number of frames in that period.

Convergence index. The convergence index was defined as the FocRtime during the second half of the trial minus the FocRtime during the first half of the trial.

Combined Success Factor (CBF). The success in a relation to the experimental limitations:

$$CBF = \frac{\text{success rate} - \text{chance level}}{\text{trial time}}$$

In our experiments the chance level was 0.2 (one out of 5 shapes or objects) divided by the trial’s time limit (in our experiments 30 or 20 seconds). CBFs are presented in **Table S1**.

Segment and SegmentTime. *Segment* was defined as the movement between two sequential direction changes. A direction change is a sharp movement change of less than $\pi/3$ in both the first and second order. *SegmentTime* is the segment duration (time in seconds from one direction change to the following).

Statistical analysis. All statistical analyses were carried out using MATLAB software. Unless mentioned otherwise statistical significance of comparisons was evaluated by two-tailed unpaired t-test, and correlations were Pearson correlations. P- values of <0.05 were considered statistically significant. Population values are described as mean \pm SEM.