

Fast, accurate reaching movements with a visual-to-auditory sensory substitution device

S. Levy-Tzedek^{a,b,*}, S. Hanassy^a, S. Abboud^a, S. Maidenbaum^a and A. Amedi^{a,b,c,*}

^a*The Institute for Medical Research Israel-Canada (IMRIC), Faculty of Medicine, Department of Medical Neurobiology, The Hebrew University of Jerusalem, Jerusalem, Israel*

^b*The Edmond and Lily Safra Center for Brain Sciences (ELSC), The Hebrew University of Jerusalem, Jerusalem, Israel*

^c*The Cognitive Science Program, The Hebrew University of Jerusalem, Jerusalem, Israel*

Abstract. Purpose: Visual sensory substitution devices (SSDs) use sound or touch to convey information that is normally perceived by vision. The primary focus of prior research using SSDs was the perceptual components of learning to use SSDs and their neural correlates. However, sensorimotor integration is critical in the effort to make SSDs relevant for everyday tasks, like grabbing a cup of coffee efficiently. The purpose of this study was to test the use of a novel visual-to-auditory SSD to guide a fast reaching movement.

Methods: Using sound, the SSD device relays location, shape and color information. Participants were asked to make fast reaching movements to targets presented by the SSD.

Results: After only a short practice session, blindfolded sighted participants performed fast and accurate movements to presented targets, which did not differ significantly from movements performed with visual feedback in terms of movement time, peak speed, and path length. A small but significant difference was found between the endpoint accuracy of movements under the two feedback conditions; remarkably, in both cases the average error was smaller than 0.5 cm.

Conclusions: Our findings combine with previous brain-imaging studies to support a theory of a modality-independent representation of spatial information. Task-specificity, rather than modality-specificity, of brain functions is crucially important for the rehabilitative use of SSDs in the blind and the visually impaired. We present the first direct comparison between movement trajectories performed with an SSD and ones performed under visual guidance. The accuracy level reached in this study demonstrates the potential applicability of using the visual-to-auditory SSD for performance of daily tasks which require fast, accurate reaching movements, and indicates a potential for rehabilitative use of the device.

Keywords: Sensory substitution, motor control, vision rehabilitation, spatial processing, visual impairment, blindness, sensorimotor integration, perception and action

1. Introduction

Sensory substitution devices (SSDs) convey information that is normally perceived by one sense, using

an alternative sense. For example, images, usually perceived with the sense of vision, can be conveyed by touch (Bach-Y-Rita et al., 1969; Bach-y-Rita, 1970; Collins, 1970) or by sound (Meijer, 1992). The ultimate goal when developing SSDs for visually impaired individuals is to assist them in perceiving the visual scene surrounding them. The ideal SSD would assist not only in sensing the environment (e.g., recognizing objects) but also in performing daily activities based on

*Corresponding authors: S. Levy-Tzedek or A. Amedi, Faculty of Medicine, The Hebrew University Building 3, 5th Floor, Room 44 P.O. Box 12272, Jerusalem 91120, Israel. Tel.: +972 2 6757259; Fax: +972 2 6758602; E-mails: shelly@huji.ac.il (S. Levy-Tzedek); amir.amed@ekmd.huji.ac.il (A. Amedi).

this input, including making accurate reaching movements toward objects, and interacting with people and objects. Potentially, it would be used for recreational activities as well, such as playing bowling. Thus, the ability to perform fast and accurate reaching movements based on information from an SSD is a crucial aspect of integration of the SSD as a useful aid to the blind and visually impaired population. However, this aspect of using SSDs has been studied to a limited extent so far.

Three of the most researched SSDs are the Tactile-Vision Sensory Substitution (TVSS), the vOICe and the Prosthesis for Substitution of Vision with Audition (PSVA), which relay information from black-and-white or grayscale images via either touch or sound. The TVSS transforms image information into tactile stimulation (either vibration or electrical stimulation) on the skin (located on the abdomen, back, thigh, fingertip, or forehead) (Bach-y-Rita and Kercel, 2003). More recently, electrical stimulation has been applied to the tongue, via a tongue display unit (TDU; (Bach-y-Rita et al., 2003). The vOICe, by contrast, converts visual information into an auditory representation termed a ‘soundscape’: the columns of an image are “sounded” sequentially from left to right, where the height of a pixel on the image is translated into sound frequency (higher frequency for higher pixels) and the amplitude of the sound corresponds to the brightness of the pixel (higher amplitude for brighter pixels) (Meijer, 1992). The PSVA also creates an auditory representation of an image, but uses a simplified dual-resolution model of the human retina. In this model, the central pixels in the image are further pixelated (for increased resolution), representing the fovea. The image is “sounded” with frequencies increasing from left to right and from bottom to top; The brightness of the pixel modulates the amplitude of the sound (Capelle et al., 1998; Renier and De Volder, 2010).

Blind users of SSDs have been successful in recognizing objects, faces (Ward and Meijer, 2010), and even specific individuals (Bach-y-Rita, 1972; p.6). Indeed, the main focus of studies using SSDs has been recognition of shapes and objects (e.g., Arno et al., 2001; Amedi et al., 2005; Renier et al., 2006; Amedi et al., 2007; Merabet et al., 2009), while relatively little attention was given to using the SSDs to guide movement. Notable exceptions are the studies by (Jansson, 1983; Auvray et al., 2007; Renier and De Volder, 2010) and (Chebat et al., 2011), which used the TVSS, the vOICe, the PSVA and the TDU, respectively, to perform a

range of movements, from a walk-and-point task to navigation down a corridor with obstacles. These studies were all important in demonstrating that movement guided by SSDs was possible. None of them, however, performed a detailed trajectory analysis or a direct comparison of these movements to ones guided by visual feedback.

Here, we use a new SSD algorithm developed in our lab, which conveys shape, location and color information using sound. The device, named the EyeMusic, represents high locations on the image as high-pitched musical notes on a pentatonic scale, and low vertical locations as low-pitched musical notes on a pentatonic scale, a scheme which has been shown to be perceived as natural (Melara and O’Brien, 1987). The EyeMusic conveys color information by using different musical instruments for each of the four colors: white, blue, red, green; Black is represented by silence. The EyeMusic currently employs an intermediate resolution of 24×40 pixels (Hanassy et al., unpublished data). Using higher resolution is possible but the unique pleasantness of this algorithm diminishes with increased resolution.

We show that with minimal practice, blindfolded sighted participants who were otherwise naïve to the use of SSDs, were able to perform fast and accurate SSD-guided reaching movements. We compare their SSD-guided movements to movements they performed with visual guidance. This study is the first, to the best of our knowledge, to perform a thorough quantitative analysis and comparison of movements under SSD guidance vs. under visual guidance.

2. Methods

2.1. Participants

18 participants (age: 24.5 ± 1.2 years; 10 females, 8 males) took part in this experiment after giving their informed consent. The participants had between 0 and 21 years of musical training. The experiment consisted of a familiarization session and a test session.

2.2. Experimental protocol

2.2.1. Familiarization session

During the familiarization session, the participants learned the basics of interpreting the visual information conveyed by sound when using the EyeMusic

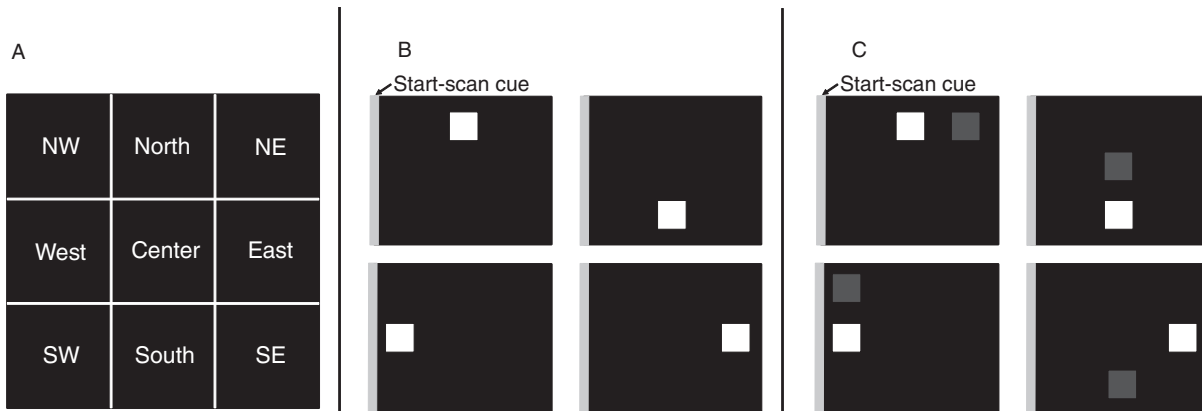


Fig. 1. The experimental display: (A) The division of the screen into a 3×3 grid (NW, NE, SW and SE stand for North West, North East, South West and South East, respectively); (B) The four possible locations of the white square (from top left, clockwise: North, South, West, East); The gray bar on the left denotes the time at the beginning of the scan when the start-scan cue was sounded. The gray bar was not shown on the screen; (C) Four examples of white-blue pairs sounded during familiarization session II (blue square appears in gray here). Here, too, the gray bar on the left is for the purpose of visualizing the start-scan cue only.

algorithm. An auditory cue (beep) was sounded at the beginning of each left-to-right scan of the image (the cue is represented by a gray bar on the left of the images in Fig. 1). It was explained to participants that: (1) the higher musical notes represent pixels that are located higher on the y-axis of an image, (2) the timing of the sound after the cue indicates the x-axis location of the pixel (that is, an object located on the left of the image will be “sounded” earlier on than an object located further on the right), and (3) different colors are represented by different musical instruments. Participants were blindfolded throughout the familiarization session, which consisted of two stages, as detailed below.

2.2.1.1. Stage I: A single object. Participants listened to a single series of 16 images. The images were of a white square against a black background, located in either the North, the South, the East or the West (see Fig. 1A, B). Each of the four images was played as a soundscape by the EyeMusic a total of four times in a randomized order. The white pixels were represented by a piano (sample sound recordings available on: http://brain.huji.ac.il/supp_mat/).

2.2.1.2. Stage II: Two different-colored objects. Participants listened to four series of 15 images each. In each of the series, the location of the white square was fixed (North/South/East/West), and a blue square was located in one of the five immediately adjacent locations (on a 3×3 grid of possible locations, see

Fig. 1A, C; for example, the white square in the North could be paired with a blue square only in one of the following locations: the NorthWest, the West, the Center, the East or the NorthEast). Each pair of white and blue squares (e.g., white in square in the South, blue square in the East) was repeated 3 times in a pseudo-randomized order within the series. The blue pixels were represented by a marimba (sample sound recordings available on: http://brain.huji.ac.il/supp_mat/).

In both stages I and II, which were performed consecutively, participants had to identify the location of the square(s) presented via the EyeMusic, and their verbal responses were recorded, and later scored. They received feedback on their responses.

2.2.2. Test session

Participants used a pen-shaped stylus to perform 2D reaching movements with their dominant hand on top of a digitizing tablet (WACOM Inc., resolution: 0.25 mm). The stylus was held by the participants' closed fist, to eliminate movement in the finger joints. Movements were made from a center location to one of four 2-cm-radius targets (North/South/East/West; see Fig. 1B), represented by a white square located 6 cm from the center. Participants performed two blocks of trials, which differed by the type of feedback provided: either visual (VIS) or auditory (SSD; via the EyeMusic). During the SSD block, participants were blindfolded; During the VIS block, the participants'

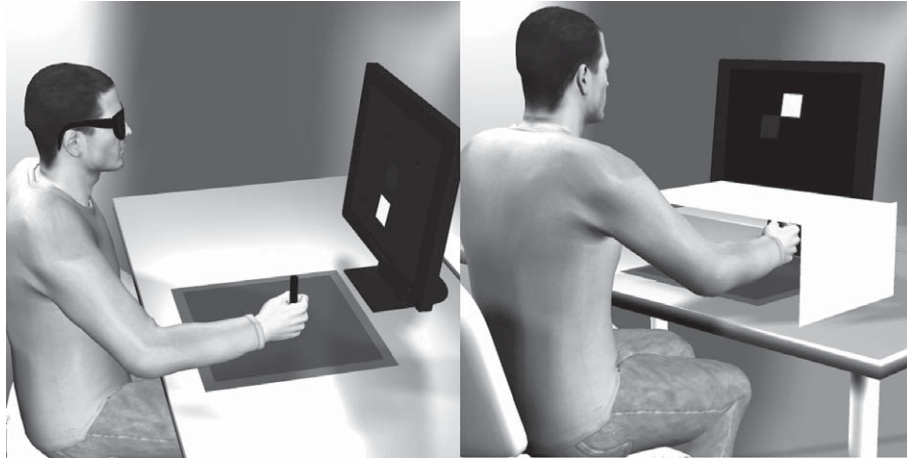


Fig. 2. The experimental setup: A participant performing: (A) The SSD block, blindfolded; (B) The VIS block, with his forearm occluded from view by an opaque cover.

arm was placed under an opaque cover, such that they did not have direct visual feedback of their hand (see Fig. 2; (Levy-Tzedek et al., 2010; Levy-Tzedek et al., 2011a)). Each block (VIS/SSD) consisted of 6 practice trials and 18 test trials, for a total of 24 trials per block (6 trials per target location, in a pseudo-random order), each lasting 1 sec. Only the test trials were subject to further analysis. Target presentation was not limited in time; rather, each target was shown (VIS) or sounded (SSD) continuously until participants were ready to perform the movement, at which point they used the hand-held stylus to tap the center of tablet, which marked the beginning of the trial, and had 1 sec to complete the movement towards the target.

Feedback was given as to the location of the target (a white square, either seen (VIS) or heard (SSD)) throughout each trial. Participants did not receive feedback on their movement path, but received feedback on the location of their hand at the end of each trial (“knowledge of results”) in the form of a blue square located at their final hand position. Feedback on the path was not provided during the movement, as we were interested only in feed-forward movements, without online feedback-based corrections. If the endpoint location was within 2 cm of the center of the target, the trial was considered successful (no error), and only feedback on the location of the *endpoint* was given (in the form of a blue square, without the target white square). Otherwise, if the end location of the hand was farther than 2 cm away from the center of the target, feedback on the location of both *target* (white) and *endpoint* (blue) was given, such that participants could use

their end position relative to the target to correct future movements. It is important to note that while during the familiarization session, the blue square could appear only in one out of five specific locations surrounding the white square on the 3×3 grid shown in Fig. 1A, the blue square could appear anywhere on the screen during the test session, limited only by the 960 pixels (24×40) resolution of the EyeMusic.

In the SSD block, participants were blindfolded, and received feedback via the EyeMusic. Participants could distinguish between the target and the endpoint in terms of color and location by noting the different instruments (piano for white and marimba for blue), their relative timing of appearance (objects on the left “sound” earlier after the start-scan cue (beep) than objects on the right), and their relative pitch (the higher the pitch of the musical note, the higher the location on the image).

Half of the participants performed the VIS block first, and the other half performed the SSD block first.

The protocol was approved by the university’s committee on human research.

2.3. Data analysis

2.3.1. Familiarization session

The verbal responses of the participants were scored as “correct” (error = 0) or “incorrect” (error = 1) per each presented object, such that during stage I of the familiarization session (a single object presented), the maximum possible error per presented image (trial) was 1, and during stage II (two objects presented), the

maximum possible error per trial was 2 (if the participant wrongly identified the location of both white and blue objects).

2.3.2. Test session

A range of 90 degrees was defined around each of the targets (45 deg to either side of a straight line connecting the center location and the target), as the “possible zone” for this target. If a participant’s movement terminated outside this range, it was eliminated from the analysis, as it was assumed that the source of such an error is not in the ability to reach the target, but rather a failure to identify the target location. This ability was separately assessed using data from the familiarization session. It is one that is expected to improve with further practice. In analyzing the data from the Test Session, however, we were mainly concerned with the characteristics of movement, and not with errors that were based on misidentification. 33 trials were excluded from a total of 324 test trials (18 trials \times 18 participants) with SSD feedback, and no trials with VIS feedback were excluded. 72% (13 out of 18) of the participants had 0–2 trials excluded.

Position and velocity traces were filtered using a first-order Butterworth filter (cutoff 20 Hz). A velocity threshold was applied to define movement onset and termination as follows: the time at which movement speed reached its peak was identified, and the velocity trace was scanned backwards until 1% of peak speed was encountered; that point was defined as the movement onset. Similarly, the velocity trace was scanned forward from the time at which movement speed peaked until 1% of peak speed was encountered; that point was defined as the time of movement termination. All analyses that follow were performed on the movements as defined from onset to termination.

2.3.2.1. Performance metrics. Reaching movements have been shown to follow a straight trajectory (Flash and Hogan, 1985), and performance measures were developed based on this characterization. Kinematic parameters such as movement time (Chen et al., 2007; Mazzoni et al., 2007), peak speed (Chen, et al., 2007), and path length to a target (Levy-Tzedek et al., 2007) have commonly been used to characterize movements. We used the following measures to characterize participants’ reaching movements under the two feedback conditions: (1) movement time: the time elapsed from movement onset to termination, (2) peak speed: the maximal hand speed during the movement, (3) path

length: the total displacement of the hand from the beginning to the end of the movement, and (4) endpoint error: the distance between the hand’s final position and the target.

2.3.2.2. Statistical analysis. A 1-way ANOVA was used to perform all statistical comparisons reported in this paper.

3. Results

3.1. Familiarization session

During stage I of the familiarization session, when participants were asked to identify the location of a single object presented via the EyeMusic, the error in their responses dropped from 0.72 ± 0.11 in the first trial to 0.0 in the last (16th) trial (mean \pm SE; see Fig. 3). This learning effect was significant ($p < 0.0005$).

When a second object was initially introduced during stage II of the familiarization session, the error in the participants’ responses increased to 0.61 ± 0.10 , but by the end of stage II dropped to 0.14 ± 0.05 . This learning effect was also significant ($p < 0.003$; see Fig. 3; For presentation purposes, the four series in stage II of the training were averaged and collapsed into a single series of 15 trials).

3.2. Test session

The average trajectories performed in the two blocks are shown in Fig. 4.

Quite surprisingly, there was no significant difference between movements performed with SSD feedback compared to those performed with visual feedback in terms of movement time, peak speed and path length (see Table 1 and Fig. 5). The endpoint error in the VIS block was slightly but significantly smaller than that in the SSD block (see Table 1 and Fig. 5). To test whether the latter is the result of signal-dependent noise (Harris and Wolpert, 1998), we compared the peak accelerations in both conditions, and found them to not differ significantly ($p = 0.12$, data not shown).

There was no effect of block order on performance.

We found no correlation between the number of years of musical training and mean error ($r^2 = 0$, polynomial and exponential fitting), and low correlation between the number of years of musical training and the number of excluded trials ($r^2 = 0.10$

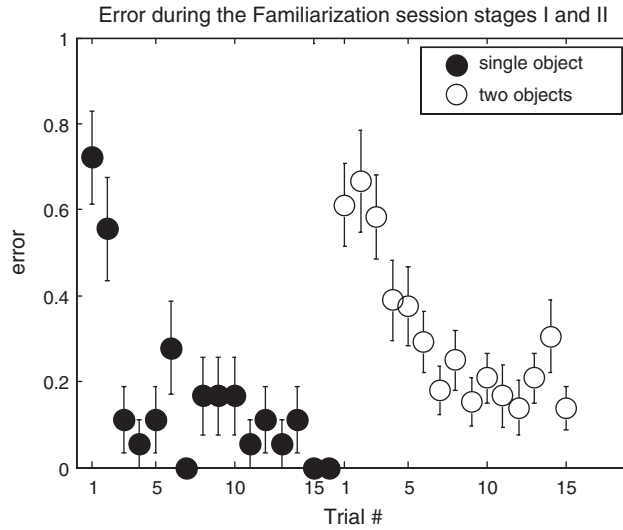


Fig. 3. Error during the familiarization session: The average error (\pm standard error) during the familiarization session, stage I (single object; black circles) and stage II (two objects; white circles).

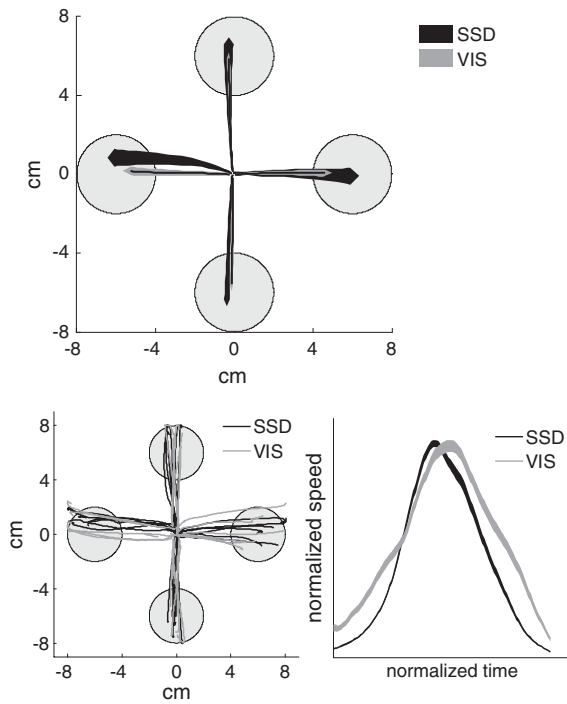


Fig. 4. Movement trajectories: Top panel: The average path (\pm standard error), across all participants, of the reaching movements performed with SSD feedback (black) and with visual feedback (gray) to the targets (gray circles). Bottom panel: Movement trajectories of a single participant (left); The average, normalized, velocity profile (\pm standard error) of a single participant (right).

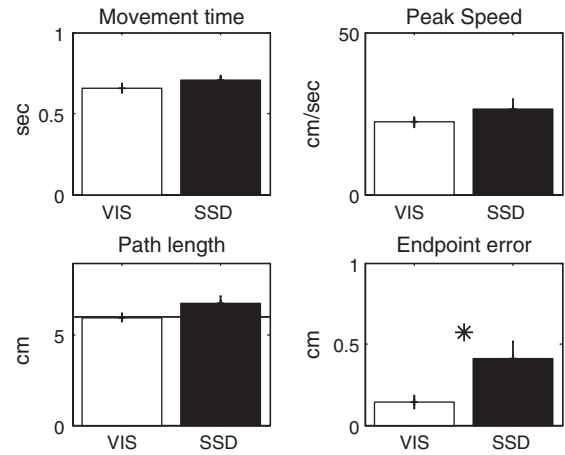


Fig. 5. Performance on the two experimental blocks (SSD/VIS): From top left, clockwise: movement time, peak speed, path length (the horizontal black bar represents the distance of the target from the center point), and endpoint error. An asterisk denotes a significant difference between the blocks, and the error bars represent standard error.

Table 1
Performance on the two experimental blocks (SSD/VIS)

	SSD (mean \pm SE)	VIS (mean \pm SE)	<i>p</i> -value
Movement time (sec)	0.71 \pm 0.02	0.66 \pm 0.03	0.17
Peak speed (cm/sec)	26.4 \pm 3.0	22.4 \pm 1.6	0.24
Path length (cm)	6.8 \pm 0.4	6.0 \pm 0.2	0.11
Endpoint error (cm)	0.41 \pm 0.1	0.14 \pm 0.04	0.02

(adjusted $r^2 = 0.05$) with a first-order polynomial fitting, $r^2 = 0.17$ (adjusted $r^2 = 0.12$) with a single-term exponential fitting).

4. Discussion

We tested the ability of naïve sighted individuals to perform movements guided by a novel sensory substitution device, and compared the quality of these movements to those performed under visual guidance. Participants underwent a short familiarization session (as short as 25 min), which involved no movement, but only listening to cues and learning to identify location and color information based on sound. The familiarization session was followed by an extremely short training session, which included movement (6 trials, 1 sec each). Participants then generated SSD-guided movements that were not different from movements made under visual guidance in terms of movement time, peak speed, and path length. Average endpoint error in both types of movements was smaller than 0.5 cm. That is, participants were able to use auditory information to create a relatively precise spatial representation. It is likely that with further practice participants can perform movements with an even smaller error using the EyeMusic SSD.

Well-trained musicians have been shown, using fMRI, to have strong audio-motor associations (Hasegawa et al., 2004; Haslinger et al., 2005; Bangert et al., 2006). The fact we found little, if any, impact of musical training on performance in the current study may be partially due to the nature of the required task, which is opposite to the tasks used in the previous studies: while in the cited studies, participants used motor actions to generate music, we had participants listen to musical notes in order to generate movement, which is not playing an instrument. Under these conditions, it is like that any such strong associations, if they existed in the participants, would not be elicited by the task after a short training period.

While the current study was a behavioral one, its results are in line with the finding, based on evidence from neural recordings, that the dorsal stream participates in mediation of sensory-motor control, even when the sensory modality is not vision (Fiehler et al., 2008), suggesting a modality-independent representation of action control. Though results from the behavioral and the imaging studies are consistent, further studies are necessary to demonstrate a dorsal pathway involvement when performing movements

based on SSD input. Indeed, it would be instructive to use a setup similar to the one we used here, in conjunction with brain-imaging techniques, to study the neural correlates of learning new sensory-motor integration rules as participants actually make movements and receive feedback on their movements via the SSD, which conveys spatial information using an auditory representation of visual information.

These results do not preclude the possibility that our participants, rather than using auditory information directly to control movement, “translated” it into a visual image of the target location, and acted based on this image. Interestingly, it has been demonstrated in several studies (e.g., Rossetti et al., 1994) that direct vision of the target immediately prior to movement significantly increased movement accuracy, whereas a delay between target presentation and movement resulted in deteriorated accuracy; that is to say, the longer participants had to use their cognitive representation of the target, the less accurate were their movements; this line of results was the basis for the suggestion that vision is needed to calibrate the proprioceptive map; and yet, participants in the current study were able to use the auditory information (possibly to create a ‘visual’-like representation which is the ultimate goal of SSDs for the blind), without any possibility to perform vision-based calibration between trials, to both plan the movements and perform corrections (on subsequent trials).

A future iteration of this experiment with congenitally blind individuals would allow us to further probe the ability to create a relatively precise spatial representation, and act on it, based on auditory information without the possibility of mediation by visual imagery.

The findings we report here demonstrate the rehabilitative potential of the EyeMusic SSD, and combine with brain-imaging studies (Renier et al., 2010; Collignon et al., 2011; Striem-Amit et al., in press) to support a theory of a modality-independent representation of spatial information. That is, the representation of spatial information may not be dependent on the modality with which the spatial information is perceived.

Our finding that movement time was not different between the two feedback types is especially important in light of previous reports that movement time increased with both increasing task difficulty and decreasing availability of visual information (Bootsma et al., 2002). Under both feedback types the movement time (~ 0.7 sec) was much shorter than the upper limit afforded by the task (1 sec).

The participants' ability to maintain the correct distance to the target during the SSD block – in other words, their ability to maintain the appropriate gain calibration of their movements based on the SSD feedback – suggests they were making use of the feedback given at the end of the trials to calibrate their movements, as opposed to performing the movements based on memory (e.g., from the VIS block), as movements in the absence of visual feedback have been shown to be miscalibrated, and not corrected based on proprioception alone (Soechting and Flanders, 1989; Wolpert et al., 1995; Levy-Tzedek et al., 2011b; Levy-Tzedek et al., 2011c).

While movement under SSD guidance has been previously demonstrated, earlier studies differed from the current one in several important respects, which we detail below, and none performed a direct comparison of the movements with those done under visual guidance.

Jansson (1983) reported that two well-trained blind individuals were able to intercept, with a one-dimensional horizontal movement of their hand, the movement of a rolling ball (the time of approach of the ball was approximately 3.5 sec); These same two individuals were also able to walk two meters towards a target, and point at it (time to complete task ~15 sec); Finally, a single well-trained blind participant was reported to slalom-walk around two vertical poles towards a third (with some collisions). These were important proof-of-principle demonstrations of the possibility to perform movement in response to information provided by an SSD (the TVSS, in this case). They were performed on 1-2 highly trained individuals, and no quantitative analysis of the movements was reported. While no direct comparisons with visually guided movements were made, it was indicated that variability of the movements was much greater than with visual guidance, and that the time to complete the walk-and-point task was considerably longer than needed under visual guidance (Jansson, 1983). In a more recent and extensive work, employing the TDU with both blind and sighted participants, Chebat et al. (2011) demonstrated that the SSD can be used to recognize the existence of obstacles (~80% success), their type (floor/side obstacles), relative size (small/large), and relative distance (near/far), and avoid some of them when walking down a corridor (~40–80% success, depending on size and type of obstacle). This study involved a large-scale whole-body movement, and information was reduced to the 10×10 resolution

of the TDU. Our current study, in contrast, involved finer upper-arm movements, guided by an SSD with a 24×40 resolution.

In Auvray et al.'s 2007 study, six blindfolded sighted participants underwent 3 hours of training before performing a walk-and-point task. They were positioned 1 m away from the target, and had to walk towards it and point at it. Participants used a hand-held camera to obtain a representation of the target in relation to their current position. Unlimited time was afforded for execution of the task, and over the course of two hours, participants shortened the time it took them to perform this task from 2 min to 1 min. This study is important in showing that an audio-based SSD (the vOICe) can be used to guide movement in a 3-D setting. We focused on a simpler reaching task, but used a time scale that is highly relevant in performing everyday tasks (movement time for natural reaching was found to be on the order of 1 sec (Messier et al., 2003)). In our experiment, the time it took participants to locate the target, reach for it, receive feedback on their actual endpoint relative to the target (and perhaps take a short break prior to initiating the next trial) was on average 26 sec. While this is an encouraging result, this interval was much shorter in the VIS block (about 6 sec on average). Though there is an important limitation of the visual-to-auditory SSDs, where the lower limit for detection of a complete scene is defined by the scan time (approximately 2 sec in the current study), we expect the latency (time to movement) to drop with practice. This expectation is based, *inter alia*, on our experience with experienced blind users of the vOICe system, who have demonstrated their ability to recognize facial expressions (a far more complex visual scene, and therefore, a more complex task in terms of perception than in the current experiment) in a matter of seconds. A video demonstration of this is presented in Striem-Amit et al.'s recent article (Striem-Amit et al., 2012; supplementary materials).

Renier and DeVolder (2010) tested sighted and blind individuals, who had extensive experience using the PSVA, on a task in which they had to identify the location of a cube in 3-D space, which was then removed by the experimenter, and had to be replaced in its original location by the participant. This study is yet another important contribution to the SSD literature, in reiterating the finding that an audio-based SSD (the PSVA) can be used to guide movement based on information collected by a head-mounted camera. In that study, time to complete the task was not reported, and mean

positioning errors ranged from 5 to 19 cm. We show that, with minimal training, participant's movements reached an accuracy level that is better than 0.5 cm.

The latter two studies demonstrate that it is possible to use a device which conveys 2D information (such as the vOICE, the PSVA, and by extension, the EyeMusic), and relay through it information on a 3D world. Indeed, it has been demonstrated that individuals were able to extract depth cues from 2D images conveyed via an SSD, such as the PSVA (see, for example, Renier et al., 2005). Undoubtedly, users of these devices would greatly benefit from the addition of input from direct distance-estimation sensors (e.g., Maidenbaum et al., 2011).

While SSDs have been around for several decades, they are not yet widely used by the visually impaired population. The reasons for this are varied, and include ease of use, and effective transmission of information. We demonstrate in this study that the EyeMusic, a novel SSD, which employs pleasant musical scales to convey visual information, can be used after a short training period to guide fast and accurate movements, similar to movements performed under visual guidance. The accuracy level reached in this study renders performing daily tasks with the aid of the SSD feasible, and indicates a potential for rehabilitative use of the device. These results have potentially important implications both theoretically, and in terms of using SSDs for 'visual' rehabilitation in the blind and in the visually impaired. Our results suggest that the representation of space might not be dependent on the modality through which the spatial information is received. This hypothesis is bolstered by brain-imaging studies indicating that areas within the visual dorsal stream, such as the precuneus (Renier, et al., 2010) and the right anterior middle occipital gyrus (MOG) (Renier, et al., 2010; Collignon, et al., 2011), showed a functional preference for spatial information delivered using the auditory (Renier, et al., 2010; Collignon, et al., 2011) or tactile modalities (Renier, et al., 2010) in congenitally blind and in sighted individuals. Additionally, we recently showed that these brain areas are also activated when sound *encodes* spatial information, rather than arrive from different spatial locations, in both sighted and congenitally blind individuals (Striem-Amit, et al., in press). A task-specific, rather than a modality-specific structure of brain areas was similarly suggested by research indicating that, during a spatial navigation task with the TDU, congenitally blind individuals were found to

recruit the posterior parahippocampus and posterior parietal and ventromedial occipito-temporal cortices, areas that are involved in spatial navigation under full vision (Kupers et al., 2010); similarly, the dorsal pathway has been shown to mediate movement control when only kinesthetic, and no visual information was used to guide actions (Fiehler et al., 2008), even when tested in congenitally blind individuals, with no prior visual experience (Fiehler et al., 2009).

It has been suggested that blind individuals tend to rely on a more egocentric, rather than an allocentric frame of reference (Röder et al., 2007; Cattaneo et al., 2008). Our current study lends support to the hypothesis that the representation of space in the brain is modality-independent, or that very little training is required to create a 'visual'-like representation of space, using sounds, to guide fast and accurate movements. This indicates a great potential in using SSDs to provide detailed spatial information for people with impaired vision, including congenitally blind individuals, allowing them to employ an allocentric spatial frame of reference when interacting with their immediate environment, and successfully make accurate movements in space based on this information.

Acknowledgements

The authors wish to thank T. Johnson for help with collecting the data, E. Striem-Amit for helpful comments on a previous version of the manuscript, and R. Leib for creating the illustrations in Fig. 2. This work was supported by the International Human Frontiers Science Program Organization Career Development Award (to AA; grant number CDA-0015/2008-070509); EU-FP7 MC International Reintegration grant (to AA; grant number MIRG-CT-2007-205357-250208); James S. McDonnell Foundation scholar award (to AA; grant number 220020284); Israel Science Foundation (grant number ISF 1684/08); The Edmond and Lily Safra Center for Brain Sciences (ELSC) Vision center grant (to AA) and an ELSC postdoctoral fellowship (to SL).

References

- Amedi, A., Kriegstein, K., Atteveldt, N.M., Beauchamp, M.S. & Naumer, M.J. (2005). Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res*, 166(3), 559-571.

- Amedi, A., Stern, W.M., Camprodon, J.A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat Neurosci*, 10(6), 687-689.
- Arno, P., Vanlierde, A., Streeel, E., Wanet Defalque, M.C., Sanabria Bohorquez, S. & Veraart, C. (2001). Auditory substitution of vision: Pattern recognition by the blind. *Appl Cognitive Psych*, 15(5), 509-519.
- Auvray, M., Hannelton, S. & O'Regan, J.K. (2007). Learning to perceive with a visuo-auditory substitution system: Localisation and object recognition with 'The vOICE'. *Perception*, 36(3), 416-430.
- Bach-y-Rita, P. (1970). Neurophysiological basis of a tactile vision-substitution system. *IEEE T Man Machine*, 108-110.
- Bach-y-Rita, P. (1972). *Brain mechanisms in sensory substitution*, Academic Press.
- Bach-Y-Rita, P., Collins, C.C., Saunders, F.A., White, B. & Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, 221(5184), 963-964.
- Bach-y-Rita, P. & Kercel, S.W. (2003). Sensory substitution and the human-machine interface. *Trends Cogn Sci*, 7(12), 541-546.
- Bach-y-Rita, P., Tyler, M.E. & Kaczmarek, K.A. (2003). Seeing with the brain. *Int j hum-comput int*, 15(2), 285-295.
- Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., et al. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *Neuroimage*, 30(3), 917-926.
- Bootsma, R.J., Boulard, M., Fernandez, L. & Mottet, D. (2002). Informational constraints in human precision aiming. *Neurosci Lett*, 333(2), 141-145.
- Capelle, C., Trullemans, C., Arno, P. & Veraart, C. (1998). A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *Biomedical Engineering, IEEE T Bio-Med Eng*, 45(10), 1279-1293.
- Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., et al. (2008). Imagery and spatial processes in blindness and visual impairment. *Neurosci Biobehav R*, 32(8), 1346-1360.
- Chebat, D.R., Schneider, F.C., Kupers, R. & Ptito, M. (2011). Navigation with a sensory substitution device in congenitally blind individuals. *Neuroreport*, 22(7), 342-347.
- Chen, Y.P., Kang, L.J., Chuang, T.Y., Doong, J.L., Lee, S.J., Tsai, M.W., et al. (2007). Use of virtual reality to improve upper-extremity control in children with cerebral palsy: A single-subject design. *Phys Ther*, 87(11), 1441.
- Collignon, O., Vandewalle, G., Voss, P., Albouy, G., Charbonneau, G., Lassonde, M., et al. (2011). Functional specialization for auditory-spatial processing in the occipital cortex of congenitally blind humans. *P Natl Acad Sci*, 108(11), 4435-4440.
- Collins, C.C. (1970). Tactile television-mechanical and electrical image projection. *IEEE T Man Machine*, 11(1), 65-71.
- Fiehler, K., Burke, M., Bien, S., Röder, B. & Rösler, F. (2009). The human dorsal action control system develops in the absence of vision. *Cereb Cortex*, 19(1), 1-12.
- Fiehler, K., Burke, M., Engel, A., Bien, S. & Rösler, F. (2008). Kinesthetic working memory and action control within the dorsal stream. *Cereb Cortex*, 18(2), 243-253.
- Flash, T. & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *J Neurosci*, 5(7), 1688-1703.
- Harris, C.M. & Wolpert, D.M. (1998). Signal-dependent noise determines motor planning. *Nature*, 394(6695), 780-784.
- Hasegawa, T., Matsuki, K.I., Ueno, T., Maeda, Y., Matsue, Y., Konishi, Y., et al. (2004). Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Cognitive Brain Res*, 20(3), 510-518.
- Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H. & Ceballos-Baumann, A.O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *J Cognitive Neurosci*, 17(2), 282-293.
- Jansson, G. (1983). Tactile guidance of movement. *Int J Neurosci*, 19(1-4), 37-46.
- Kupers, R., Chebat, D.R., Madsen, K.H., Paulson, O.B. & Ptito, M. (2010). Neural correlates of virtual route recognition in congenital blindness. *P Natl Acad Sci*, 107(28), 12716-12721.
- Levy-Tzedek, S., Ben Tov, M. & Karniel, A. (2011). Early switching between movement types: Indication of predictive control? *Brain Res Bull*, 85(5), 283-288.
- Levy-Tzedek, S., Ben Tov, M. & Karniel, A. (2011). Rhythmic movements are larger, faster but with the same frequency upon removal of visual feedback. *J Neurophysiol*, 106(5), 2120-2126.
- Levy-Tzedek, S., Krebs, H., Song, D., Hogan, N. & Poizner, H. (2010). Non-monotonicity on a spatio-temporally defined cyclic task: Evidence of two movement types? *Exp Brain Res*, 202(4), 733-746.
- Levy-Tzedek, S., Krebs, H.I., Arle, J.E., Shils, J.L. & Poizner, H. (2011). Rhythmic movement in Parkinson's disease: Effects of visual feedback and medication state. *Exp Brain Res*, 211(2), 277-286.
- Levy-Tzedek, S., Krebs, H.I., Shils, J.L., Apetauerova, D. & Arle, J.E. (2007). Parkinson's disease: A motor control study using a wrist robot. *Adv Robotics*, 21(10), 1201-1213.
- Maidenbaum, S., Levy-Tzedek, S., Arbel, R., Hanassy, S. & Amedi, A. (2011). *Shapes from the Depths – Shape Recognition Using Distance Information*. Presented at the 7th Computational Motor Control Workshop, Israel.
- Mazzoni, P., Hristova, A. & Krakauer, J.W. (2007). Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. *J Neurosci*, 27(27), 7105.
- Meijer, P.B.L. (1992). An experimental system for auditory image representations. *IEEE T Bio-Med Eng*, 39(2), 112-121.
- Melara, R.D. & O'Brien, T.P. (1987). Interaction between synesthetically corresponding dimensions. *J Exp Psychol Gen*, 116(4), 323.
- Merabet, L.B., Battelli, L., Obretenova, S., Maguire, S., Meijer, P. & Pascual-Leone, A. (2009). Functional recruitment of visual cortex for sound encoded object identification in the blind. *Neuroreport*, 20(2), 132-138.

- Messier, J., Adamovich, S., Berkinblit, M., Tunik, E. & Poizner, H. (2003). Influence of movement speed on accuracy and coordination of reaching movements to memorized targets in three-dimensional space in a deafferented subject. *Exp Brain Res*, 150(4), 399-416.
- Renier, L., Bruyer, R. & De Volder, A.G. (2006). Vertical-horizontal illusion present for sighted but not early blind humans using auditory substitution of vision. *Atten Percept Psycho*, 68(4), 535-542.
- Renier, L., Collignon, O., Poirier, C., Tranduy, D., Vanlierde, A., Bol, A. et al. (2005). Cross-modal activation of visual cortex during depth perception using auditory substitution of vision. *Neuroimage*, 26(2), 573-580.
- Renier, L. & De Volder, A.G. (2010). Vision substitution and depth perception: Early blind subjects experience visual perspective through their ears. *Disabil Rehabil: Assistive Technology*, 5(3), 175-183.
- Renier, L.A., Anurova, I., De Volder, A.G., Carlson, S., VanMeter, J. & Rauschecker, J.P. (2010). Preserved functional specialization for spatial processing in the middle occipital gyrus of the early blind. *Neuron*, 68(1), 138-148.
- Röder, B., Kusmirek, A., Spence, C. & Schicke, T. (2007). Developmental vision determines the reference frame for the multi-sensory control of action. *P Natl Acad Sci*, 104(11), 4753-4758.
- Rossetti, Y., Stelmach, G., Desmurget, M., Prablanc, C. & Jeannerod, M. (1994). The effect of viewing the static hand prior to movement onset on pointing kinematics and variability. *Exp Brain Res*, 101(2), 323-330.
- Soechting, J. & Flanders, M. (1989). Sensorimotor representations for pointing to targets in three-dimensional space. *J Neurophysiol*, 62(2), 582-594.
- Striem-Amit, E., Dakwar, O., Reich, L. & Amedi, A. (2011). The large-scale organization of "visual" streams emerges without visual experience. *Cereb Cortex*, doi:10.1093/cercor/bhr253
- Striem-Amit, E., Guendelman, M. & Amedi, A. (2012). 'Visual' acuity of the congenitally blind using visual-to-auditory sensory substitution. *PLoS ONE*, 7(3), e33136.
- Ward, J. & Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device. *Conscious Cogn*, 19(1), 492-500.
- Wolpert, D., Ghahramani, Z. & Jordan, M. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880-1882.