frontiers in Behavioral Neuroscience

ORIGINAL RESEARCH
published: 21 April 2016
doi: 10.3389/fnbeh.2016.00079

Sensory Substitution: The Spatial Updating of Auditory Scenes "Mimics" the Spatial Updating of Visual Scenes

Achille Pasqualotto[1]* *and Tayfun Esenkaya*[1,2]

[1] Faculty of Arts and Social Sciences, Sabanci University, Istanbul, Turkey, [2] Department of Psychology, University of Bath, Bath, UK

Visual-to-auditory sensory substitution is used to convey visual information through audition, and it was initially created to compensate for blindness; it consists of software converting the *visual* images captured by a video-camera into the equivalent *auditory* images, or "soundscapes". Here, it was used by blindfolded sighted participants to learn the spatial position of simple shapes depicted in images arranged on the floor. Very few studies have used sensory substitution to investigate *spatial* representation, while it has been widely used to investigate object *recognition*. Additionally, with sensory substitution we could study the performance of participants *actively* exploring the environment through audition, rather than *passively* localizing sound sources. Blindfolded participants *egocentrically* learnt the position of six images by using sensory substitution and then a judgment of relative direction task (JRD) was used to determine how this scene was represented. This task consists of *imagining* being in a given location, oriented in a given direction, and pointing towards the required image. Before performing the JRD task, participants explored a map that provided *allocentric* information about the scene. Although spatial exploration was egocentric, surprisingly we found that performance in the JRD task was better for allocentric perspectives. This suggests that the egocentric representation of the scene was *updated*. This result is in line with previous studies using visual and somatosensory scenes, thus supporting the notion that different sensory modalities produce equivalent spatial representation(s). Moreover, our results have practical implications to improve training methods with sensory substitution devices (SSD).

Keywords: sensory substitution, vOICe, allocentric, egocentric, perspective-taking

OPEN ACCESS

Edited by:
Nuno Sousa,
University of Minho, Portugal

Reviewed by:
Volker Korz,
Medical University Vienna, Austria
Monica Gori,
Istituto Italiano di Tecnologia, Italy
Noelle Stiles,
California Institute of Technology, USA

***Correspondence:**
Achille Pasqualotto
achille@sabanciuniv.edu

Received: *04 December 2015*
Accepted: *08 April 2016*
Published: *21 April 2016*

Citation:
Pasqualotto A and Esenkaya T (2016) Sensory Substitution: The Spatial Updating of Auditory Scenes "Mimics" the Spatial Updating of Visual Scenes. Front. Behav. Neurosci. 10:79. doi: 10.3389/fnbeh.2016.00079

INTRODUCTION

Knowing the location of external objects is critical for survival, and in many species this function depends on vision. In case of visual loss, participants involved in spatial tasks showed that the non-visual modalities could partially (but not fully) compensate for the lack of visual input (Putzar et al., 2007; King, 2009; Gori et al., 2010, 2014; Papadopoulos et al., 2012; Pasqualotto and Proulx, 2012). Among the tools created to compensate for visual loss, sensory substitution has provided excellent practical and theoretical results (Bach-y-Rita et al., 1969; Bach-y-Rita, 1972; Proulx et al., 2014b).

For example, it shed light on the mechanisms of neural plasticity in both visually impaired *and* sighted individuals (Rauschecker, 1995; Sampaio et al., 2001; Bach-y-Rita and Kercel, 2003; Tyler et al., 2003; Amedi et al., 2007; Proulx et al., 2014a; Brown et al., 2015). Sensory substitution devices (SSD) convey visual information through other modalities, thus allowing for recognition of distant and silent objects. In visual-to-auditory sensory substitution, software called "The vOICe" converts the images[1] captured by a video-camera (controlled by the user) into equivalent "soundscapes" (or auditory images), which is listened through headphones. To transform visual images into auditory images, the software scans visual images from left-to-right, converts them into grayscale images, and subdivides them into pixels; each pixel is then converted into sound (or "sonified") based on its luminance, horizontal position, and vertical position. High luminance pixels will sound *louder* than low luminance pixels, pixels on the left will be played *before* than those on the right, and pixels at the top will have a higher *pitch* than those at the bottom. Thus, visual images are sonified by using three parameters: loudness, time, and pitch (Meijer, 1992). Another feature of sensory substitution is that the active movement of the camera performed by the users of a SSD is crucial for the successful recognition of the objects (White et al., 1970; Bach-y-Rita, 1972; Lenay et al., 2003). Thus, in our study participants explored the environment by using actively moving the video-camera.

There are many studies using sensory substitution that are concerned with *recognition* tasks (i.e., "what" tasks; Ptito et al., 2005; Kim and Zatorre, 2008; Brown et al., 2011; Striem-Amit et al., 2012; Haigh et al., 2013). However, studies using The vOICe in *spatial* tasks (i.e., "where" tasks) remain sparse, and more research is needed in order to improve training regimens with SSD (Auvray et al., 2007; Proulx et al., 2008; Chebat et al., 2011). As a matter of fact, lengthy and frustrating training with SSD has been blamed for the scarce utilization of these tools (included The vOICe) in everyday life (Loomis, 2010; Maidenbaum et al., 2014). Although in some cases the input from SSD can be successfully interpreted by naïve users (Auvray et al., 2005; Stiles et al., 2015), usually training is required to perform most tasks employing SSD (Bach-y-Rita, 1972; Meijer, 1992). In the present study, before the main experiment, participants were trained to recognize simple images by using The vOICe.

Unlike the studies investigating auditory spatial representation where participants *passively* listened to sounds delivered by different sources (e.g., Klatzky et al., 2003), using The vOICe allowed for studying *active* spatial exploration (Auvray et al., 2005; Stiles et al., 2015). Therefore, participants actively explored the environment, which included the target images and environmental features such as the floor, thus providing novel insights on active spatial exploration performed through audition (Klatzky et al., 1998; Gaunet et al., 2001). Finally, the use of a visual-to-auditory sensory substitution device was necessary because audition is ill-suited for exploring silent objects (Yamamoto and Shelton, 2009; Avraamides and Kelly, 2010).

In this study we investigated how an auditory-learnt scene (multiple objects) was represented. There are two major manners to represent spatial information; egocentrically, where the spatial relations between the observer's position and the position of each object are stored in spatial memory; or allocentrically, where the spatial relations among the observed objects are stored in spatial memory (McNamara, 2003). Our purpose is to use The vOICe (i.e., auditory input) to investigate how a regularly arranged scene ("chessboard-like" arrangement) is stored in spatial memory. Previous studies using vision demonstrated that, rather than being represented according to the egocentric "viewpoint"[2] (Mou and McNamara, 2002), regularly arranged scenes were represented according to the reference frame used during scene learning. In fact, when observers learnt the scene egocentrically, the resulting spatial representation was egocentric; contrarily, when observers learnt the scene allocentrically, the resulting spatial representation was allocentric (see also Wolbers and Büchel, 2005; Pasqualotto and Proulx, 2013; Pasqualotto et al., 2013a; Thibault et al., 2013). Additionally, once spatial representation is formed, it can be *updated* by subsequent input (Simons and Wang, 1998; Mou et al., 2004). Spatial updating has been extensively studied for visually-learnt (Diwadkar and McNamara, 1997; Simons and Wang, 1998; Zhao et al., 2007) and haptically-learnt scenes (Newell et al., 2005; Pasqualotto et al., 2005). However, there is little research on auditory-learnt spatial updating (Loomis et al., 2002; Klatzky et al., 2003).

The method used in this article will be based on the study by Pasqualotto et al. (2013b), where groups of participants with different levels of visual experience (but here we will focus on blindfolded sighted) egocentrically learnt a regularly arranged scene through somatosensation; they repeatedly walked from the "viewpoint" to each object composing the scene. After the egocentric learning, participants received allocentric information about the scene (a map disclosing the regular structure of the scene, but *not* the actual objects) to investigate whether spatial updating could take place. The resulting spatial representation was investigated by a perspective-taking task (or JRD). For those unfamiliar with this task, it involves aligning themselves with an imagery perspective and pointing to the required object/landmark (Mou and McNamara, 2002; McNamara, 2003). For example, imagine that you are in your kitchen near the fridge, that you are looking towards the sink (imaginary perspective), and that you have to point towards the kitchen door (required object). Pasqualotto et al. (2013b) found that the pointing performance of blindfolded sighted participants showed the characteristic saw-tooth profile (e.g., Diwadkar and McNamara, 1997; Shelton and McNamara, 2001; Mou and McNamara, 2002), where allocentric perspectives were better performed than egocentric ones, thus suggesting that

[1]Throughout the manuscript, the term "image" will be used for both visual and sonified images and the context will clarify its exact meaning.

[2]Although the terms "viewpoint", "observer", etc. are connected to the visual modality, for sake of consistency and clarity we will continue to use them when scenes were explored through non-visual modalities.

spatial updating occurred (i.e., the egocentric representation of the scene was updated into an allocentric representation). In particular, Mou and McNamara (2002) found that the representation of a regularly arranged scene was based on the allocentric structure of the scene (i.e., spatial relations among objects) rather than on the experienced view (egocentric). In place of somatosensation, here blindfolded sighted participants used audition (i.e., the SSD) to learn the spatial location of six images.

In case the present study will replicate the results by Pasqualotto et al. (2013b), this would suggest that auditory-learnt scenes are represented in an equivalent manner to those learnt by vision and somatosensation (Giudice et al., 2011; Loomis et al., 2013; Intraub et al., 2015). Taking into consideration that, independently from the sensory modality, representing the space is subserved by the same brain areas (Kandel et al., 2012), we expect to replicate the findings by Pasqualotto et al. (2013b).

MATERIALS AND METHODS

Participants

Eighteen sighted participants (nine male) recruited among the students of the Sabanci University participated in the experiment. Their average age was 22.4 years. No participant suffered from hearing/motor impairments and all signed the informed consent form. This study was carried out in accordance with the recommendations of Declaration of Helsinki and the protocol was approved by the Sabanci University Research Ethics Committee. Participants received meal-vouchers for their participation. Each participant went to the lab three times across three consecutive days (once per day).

Apparatus

Apparatus Training Sessions

During the 2 days preceding the main experiment, participants familiarized with the six images that were going to be used during the main experiment (a triangle, a star, a moon crescent, a rectangle, a circle, and an upward bar). Although there is evidence that untrained participants can recognize sonified images (Auvray et al., 2005; Stiles et al., 2015), the use of SSD usually requires some amount of training (Bach-y-Rita, 1972; Meijer, 1992). In fact, without training our participants would have been completely confused and helpless. Microsoft™ PowerPoint presentations were used to train participants. For sake of clarity and simplicity, during the first training day the six images were presented upright and on a white background (see the top part of **Figure 1**). During the second training day the same six images were presented as they were going to be "seen"

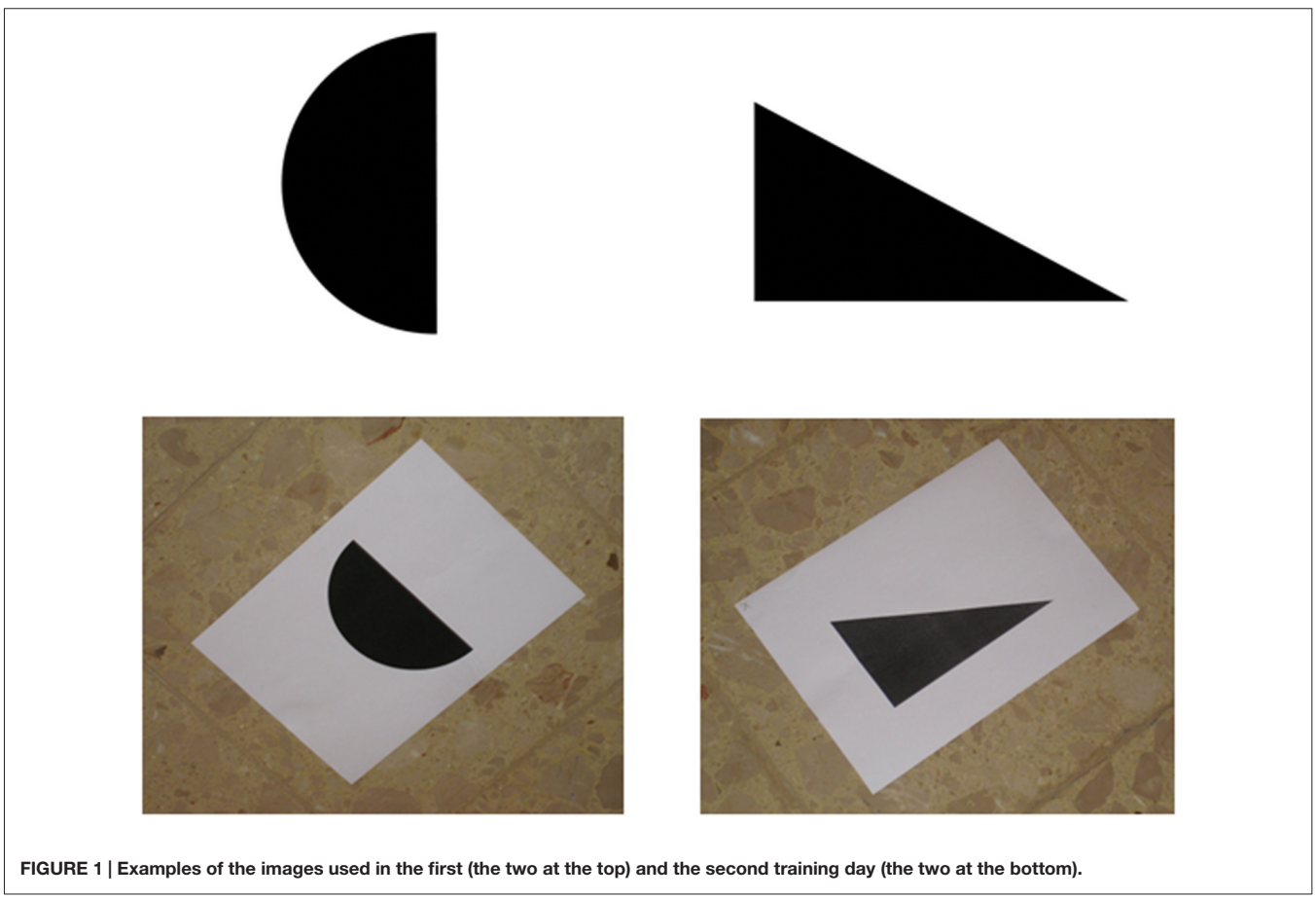

FIGURE 1 | Examples of the images used in the first (the two at the top) and the second training day (the two at the bottom).

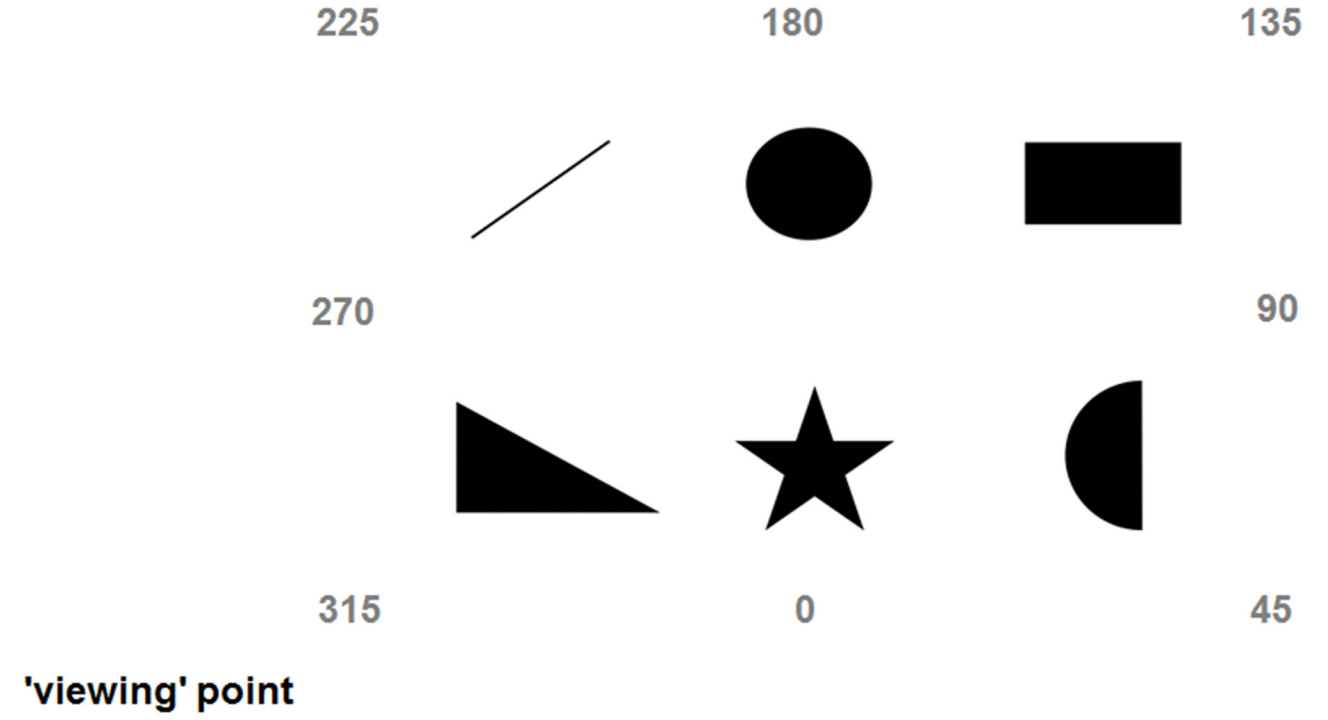

FIGURE 2 | Depiction of the scene explored during the third day (main experiment); numbers indicate the imaginary headings/perspectives (315, 45, 135, and 225 are egocentric, while 0, 90, 180, and 270 are allocentric).

during the main experiment, that is, printed on a white sheet laying on a "patterned floor" (i.e., the floor of the room where the main experiment took place). In fact, the soundscape generated by The vOICe includes all the items that in a given moment are captured by the video-camera, including the target image, the sheet, and the surrounding floor. Hence, we trained participants to recognize sonified images embedded in the "noise" produced by the sonified sheet and floor (see the bottom part of **Figure 1**). Additionally, in the second training day images were presented as participants were going to "see" them from the "viewing point" during the main experiment (i.e., tilted, see the bottom part of **Figure 1**). To better understand this point, you can look at **Figure 2**, imagine standing at the viewing point, and realize that from there images are seen as in the bottom part of **Figure 1**.

Apparatus Main Experiment

During the main experiment the six familiar images were arranged on the floor of a large room (about 5 m by 4 m), with 60 cm distance between any two of them (see **Figure 2**). Each image was printed in black-and-white on A4 sheets. The settings of The vOICe (freely available at: www.seeingwithsound.com) were: standard view, ×2 zoom, and medium loudness. The perspective-taking task was conducted in a different room (walking there took about 2 min) using a LogiTech™ 3DPro Joystick connected to a HP™ desktop running a MatLab™ program. There were 40 trials where participants *imagined* aligning themselves to eight perspectives. Four of these imaginary perspectives (20 trials in total) were called "egocentric" and consisted of the subjective perspective of the scene (i.e., the "view" from the "viewing" point, 315°), its mirror perspective (i.e., the "view" from the opposite side of the scene, 135°) and the two intermediate perspectives (45° and 225°). For example, imagining being near the triangle and facing the circle would be a 315° perspective, while imagining being near the circle and facing the triangle would be a 135° perspective (see **Figure 2**). The remaining four imaginary perspectives (20 trials in total) were called "allocentric" and were the perspectives aligned with the intrinsic axes of the scene (0°, 90°, 180°, and 270°). For example, imagining being near the triangle and facing the upward bar would be a 0° perspective, while imagining being near the upward bar and facing the triangle would be a 180° perspective (see **Figure 2**; Pasqualotto et al., 2013b). In sum, egocentric perspectives were those spatially related with the subjective view of the scene, while allocentric perspectives were those spatially related with the intrinsic axes of the scene. For sake of consistency and clarity, the labels associated to the perspectives were the same as Mou and McNamara (2002) and Pasqualotto et al. (2013b); that is, we could have labeled the subjective perspective experienced by the participants "0°" rather than "315°" and have all the other perspectives renamed (or even Ego1, Allo1, Ego2, Allo2, etc.), but we preferred to use the same names for easier comparison. During the perspective-taking task participants imagined being aligned to one of the eight perspectives and had to point to the required object (e.g., "Imagine that you are near the

triangle and that you are facing the upward bar, point to the star"; pointing the joystick to create a 90° angle would be the best answer). Therefore, the way participants performed this perspective-taking task would inform us about which spatial representation of the scene (egocentric/allocentric) they possessed.

Procedure

Procedure Training Sessions

Training was necessary so that, during the main experiment, participants would have been able to actively explore a meaningful scene. In other words, participants would have been able to recognize the images and form a spatial representation of the scene. On the first training day participants signed the consent form and familiarized themselves with the six images that were used in the main experiment. Participants were initially presented with each visual image coupled with its soundscapes (e.g., a triangle coupled with the soundscape of that triangle). Once participants were sufficiently confident (i.e., after about 3–4 repetitions), they were presented with the soundscapes alone, and were asked to declare to which visual image they corresponded (e.g., when the soundscape of a triangle was played then participants were expected to answer: "Triangle!"). The training terminated after two consecutive error-free runs (on average after 5–6 runs). The first training session lasted for about 20 min.

The day after, participants underwent the second training session. Here each image was presented (both visually and auditorily) printed on a white sheet and with the floor of the room where the main experiment was going to take place. Additionally, images were presented tilted because this is how they were going to be seen from the "viewing point" (as abovementioned, see bottom part of **Figures 1**, **2** to understand this point). This training was necessary to ensure that, during the main experiment, participants would have been able to recognize the images by audition alone. The second training session proceeded as the first one; presentation and testing that ended after two consecutive runs without errors (on average after 4–5 runs). Finally, in this session participants familiarized themselves with the use of the joystick for the JRD task. Here, they were asked to use the joystick to point towards well-known locations inside the campus (e.g., "Imagine that you are at the main gate, that you are facing the library, point to the bus station"). During the main experiment, the JRD was performed by using the six objects. The second training session lasted about 25 min.

Procedure Main Experiment

On the third consecutive day, participants run the main experiment; they were blindfolded and guided to the room where the six printed images were set on the floor (see **Figure 2**). Blindfolded participants were asked to wear headphones to listen to the soundscapes generated by The vOICe. They were instructed to stand still and were oriented along the 315° viewpoint (see **Figure 2**). To familiarize blindfolded participants with the place where they were, initially they were instructed to use the video-camera to explore the environment itself (walls, floor[3], ceiling, doors, etc.), and only then the six images arranged on the floor. Blindfolded participants were guided twice by the experimenter to each image by following the serial learning sequence: triangle, star, moon, rectangle, circle, and upward bar. This procedure was aimed to trigger the use of an egocentric reference frame (Pasqualotto et al., 2013b) and, by using a ×2 zoom, we ensured that images were "seen" one-by-one only. Thus, information on the spatial relations *among* the images was not accessible. In fact, this procedure emphasized the spatial relationships among each image and the "observer" (i.e., egocentric scene representation). During this guided learning phase, blindfolded participants had to correctly name each image. Then for about 2 min participants explored the images without guidance by following the same sequence and by naming each object.

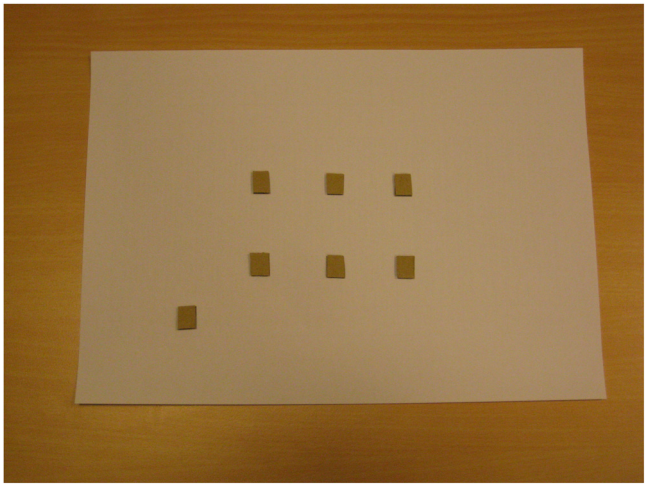

FIGURE 3 | The "map" explored by blindfolded participants; it "shows" (no vision involved) the regular structure of the scene without disclosing the identity of the images (triangle, circle, etc.). The lower-left dot represents the point where participants stood.

Subsequently, blindfolded participants were brought to the room where the perspective-taking task was performed. Before starting the task, blindfolded participants explored a map showing a bird-eye view of the regularly arranged scene (six dots), in addition to a seventh dot representing the point where blindfolded participants stood during scene exploration ("viewing" point). Blindfolded participants were told that the map represented the scene and that each dot represented one image (the identity of the images was not disclosed); in fact, the map reported seven identical raised dots (see **Figure 3**). By providing allocentric (or configurational) information, this procedure was aimed to trigger the spatial updating of the scene (see Pasqualotto et al., 2013b). Blindfolded participants used both hands to explore the map; initially they were assisted by the experimenter and then they explored the map on their own. Map exploration with the experimenter (1 min) proceeded along the same serial learning sequence followed during the exploration

[3]Not the area of the floor where images were located.

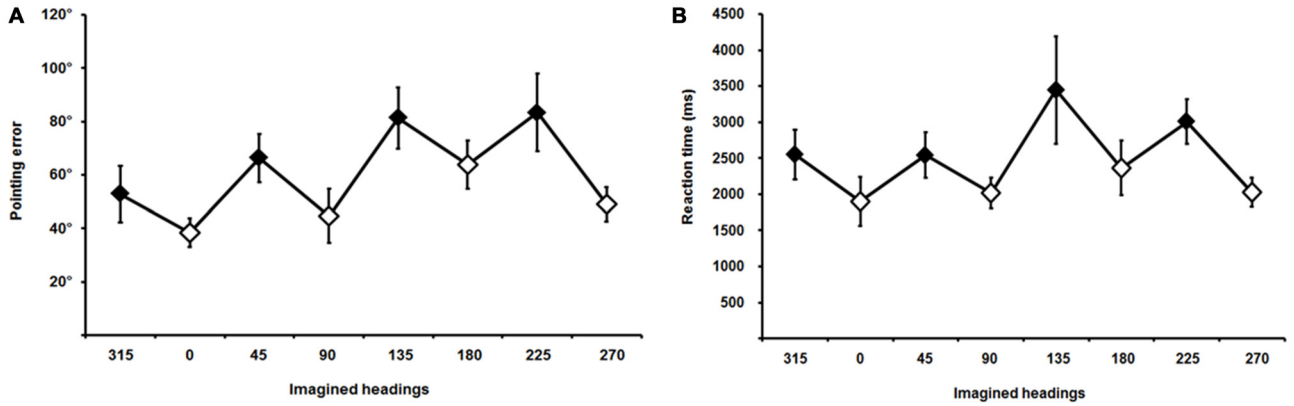

FIGURE 4 | Pointing errors (A) and reaction times (B) across different imaginary headings or perspectives; 315, 45, 135, and 225 (filled diamonds) are the egocentric and 0, 90, 180, and 270 (empty diamonds) are the allocentric perspectives. Error bars represent the ±SE.

of the *real* scene, while unassisted exploration (1 min) was unconstrained.

The perspective-taking task consisted of 40 randomized trials, 20 requiring blindfolded participants to imagine perspectives related with the subjective "view" of the scene (egocentric) and 20 related with the intrinsic axes of the scene (allocentric). The experimenter read out each trial for the blindfolded participants, for example: "Imagine that you are near the star (3-s pause), facing the rectangle (3-s pause), point to the circle!" or "Imagine that you are near the rectangle (3-s pause), facing the circle (3-s pause), point to the moon!" (see Pasqualotto et al., 2013b). An auditory cue (a "bling" sound) was played to prompt blindfolded participants' responses (i.e., joystick aiming). In the end, participants took off the blindfold and were debriefed. Pointing errors in degrees (°) and reaction times (ms) were recorded. The main experiment took about 40 min.

RESULTS

Average pointing errors and reaction times are plotted in **Figure 4**; they showed the classic saw-tooth pattern (Mou and McNamara, 2002; Pasqualotto et al., 2013b), indicating that trials involving allocentric perspectives were performed more accurately (smaller pointing errors) and more rapidly (shorter reaction times) than trials involving egocentric perspectives. Pointing errors were normally distributed[4], thus we started by analyzing their main effect across the eight imaginary perspectives (average pointing errors for each perspective and for each participant) by using one-way ANOVA; results showed that existed significant differences across the perspectives [$F_{(1,17)} = 5.41$, $p = 0.001$].

Thus, we continued the analysis with a paired-samples *t*-test comparing average pointing errors for egocentric vs. allocentric perspectives for each participant, which showed a significant effect [$t_{(17)} = -4.04$, $p = 0.001$] indicating that participants' performance in the JRD task was more accurate when the imaginary perspectives were aligned with the allocentric axes of the scene (0°, 90°, 180°, and 270°) than with the egocentric perspectives of the scene (45°, 135°, 225°, and 315°). Respectively, average pointing errors were 49.04° (standard deviation 10.82) and 71.03° (SD 14.26; see **Figure 4A**).

Reaction times were normally distributed[5] and, as we did for pointing errors, average reaction times across the eight imaginary perspectives were initially analyzed by using one-way ANOVA; results supported the existence of a significant difference across perspectives [$F_{(1,17)} = 2.93$, $p = 0.01$].

The paired-samples *t*-test performed on the average reaction times showed a significant effect of the imaginary perspective [$t_{(17)} = 2.84$, $p = 0.01$], indicating that participants pointed to the targets more rapidly when imaginary perspectives were aligned with the allocentric axes of the scene (average reaction time 2078 ms, SD 1235) than when aligned with the egocentric views (average reaction time 2886 ms, SD 1980; see **Figure 4B**).

DISCUSSION

Participants actively explored the environment by using The vOICe (audition) to learn the position of six images (a triangle, a circle, etc.); unlike experiments using vision (e.g., Mou and McNamara, 2002), before the main experiment it was useful familiarize participants with the soundscapes of a triangle, a circle, etc. After the training, in the main experiment participants used The vOICe to explore the scene in an egocentric manner, thus emphasizing spatial relationships among the "observer" and

[4]For the egocentric condition skewness was 0.048 (SE = 0.536) and a kurtosis was −1.736 (SE = 1.038); for the allocentric condition skewness was 0.820 (SE = 0.536) and kurtosis was −0.08 (SE = 1.038; Cramer, 1998).

[5]For the egocentric condition skewness was 1.286 (SE = 0.536) and a kurtosis was 2.035 (SE = 1.038); for the allocentric condition skewness was 1.120 (SE = 0.536) and kurtosis was 1.910 (SE = 1.038).

the images. Then a map was provided to investigate whether it could trigger spatial updating as in Pasqualotto et al. (2013b). Even though participants egocentrically learnt the locations of the images, the results of the JRD task suggested that the scene was allocentrically represented (Pasqualotto et al., 2013b). This suggests that the egocentric auditory scene was *updated* into an allocentric representation by somatosensory information (map), like visual scenes were updated by somatosensory information generated by self-motion (Farrell and Thomson, 1998; Burgess et al., 2004). Spatial updating has been extensively studied for single objects (Woods and Newell, 2004; Newell et al., 2005) as well as for multiple objects (Diwadkar and McNamara, 1997; Simons and Wang, 1998; Waller et al., 2002; Mou et al., 2004); visual and somatosensory modalities were also investigated (Pasqualotto et al., 2005; Mou et al., 2006; Zhao et al., 2007). Since both vision and somatosensation are well-suited to convey information about shape and the position of objects (at least within the peripersonal space), they have received substantial attention (Ballesteros et al., 1998; Kappers and Koenderink, 1999). Contrarily, spatial updating in audition is little studied because audition is well-suited for localizing objects that emit sounds (see Ho and Spence, 2005; Yamamoto and Shelton, 2009), but not for silent objects (i.e., the vast majority of the objects). We overcame this limit by employing The vOICe, which uses audition to convey information about the shape and the location of objects.

Our results showed that spatial updating occurs for actively learnt auditory scenes and, combined with previous findings on vision and somatosensation (Ballesteros et al., 1998; Avraamides et al., 2004; Lacey et al., 2007; Giudice et al., 2011), suggest that spatial representation is independent from the sensory modality used to explore the space. In other words, our results suggest that visual, somatosensory and auditory spatial information generates equivalent spatial representations. These findings are corroborated by studies showing how spatial information conveyed by different modalities is processed in the same brain areas; one of them is the posterior parietal cortex (PPC, Sakata and Kusunoki, 1992; Knudsen and Brainard, 1995; Farrell and Robertson, 2000; Makin et al., 2007; Morris et al., 2007). In fact, it has been suggested that visual, auditory, and somatosensory information is initially processed by the respective primary sensory cortices (e.g., visual information is processed by primary visual cortex) before being conveyed to "higher level" cortices via two different pathways specialized for identity and location of objects (Mishkin et al., 1983; Lomber and Malhotra, 2008). For all sensory modalities, the pathways specialized for object localization reach the posterior parietal cortex, which processes spatial information disregarding the modality that generated it (Anderson, 2010; Kandel et al., 2012). Areas processing spatial information arising from different sensory modalities include also the prefrontal cortex and the hippocampus (Ghazanfar and Schroeder, 2006; Avenanti et al., 2012; Hartley et al., 2014), and in recent times it has been found that multisensory processing occurs also in areas believed to be strictly unisensory, such as primary sensory cortices (Pascual-Leone and Hamilton, 2001; Sadato et al., 2007; Beer et al., 2011; Pasqualotto et al., 2015). Although speculative, this neuroscientific evidence can explain the results we obtained in our experiment.

An alternative explanation of our results is that, although vision was not involved, our participants created mental images of the scene. There are numerous empirical findings showing that mental imagery is largely visually based (Arditi et al., 1988; De Volder et al., 2001; Tokumaru et al., 2003; Ganis et al., 2004), therefore it is possible that our participants created visual images of the unseen scene. Recoding non-visual spatial information into visual images might explain the finding of the current study (using audition) and of other studies using non-visual modalities for spatial exploration (Yamamoto and Shelton, 2009; Avraamides and Kelly, 2010; Schifferstein et al., 2010). The role of visual mental imagery could be tested in future experiments where individuals without visual experience, and thus without visual imagery, are tested (i.e., congenitally blind participants; for a review see Pasqualotto and Proulx, 2012). Another idea for future studies is to employ The vOICe for allocentric spatial learning rather than egocentric. In the present study we used egocentric learning (images were learnt one-by-one by following a sequence), but in principle it is possible to use the sensory substitution device to explore the *entire* scene (and not image-by-image), thus emphasizing the spatial relations among objects (i.e., allocentrically). As a matter of fact, in our lab we have already started working on the latter idea; preliminary and possibly "temporary" results are suggesting that participants can learn auditory scenes (through The vOICe) with a level of accuracy comparable to visual scenes.

It is particularly important to note that in a previous study using the same methods, but involving somatosensation (Pasqualotto et al., 2013b), blindfolded sighted participants produced a much poorer performance than in the present study. Although the pattern of the results was equivalent (i.e., allocentric representation displayed by a saw-tooth pattern), the overall performance was more accurate in this study using sensory substitution. This supports the ability of The vOICe to successfully convey spatial information. Our results showed that information acquired though visual-to-auditory sensory substitution (The vOICe) can be updated by when allocentric information is provided; this finding has practical implications, because training regimens with sensory substitution could be improved by providing allocentric spatial information—as we did by using a map. Problems connected to long trainings with SSD, or the absence of training protocols, have been identified as major obstacle for the use of these tools in real-world settings (Loomis, 2010; Maidenbaum et al., 2014).

Our participants were able to update the representation of the scene from egocentric to allocentric (i.e., achieve a more "global" representation). Yet, this finding needs to be confirmed by testing visually impaired individuals. In fact, there is convincing evidence that congenitally blind (individuals with no visual experience) find particular problematic to achieve allocentric spatial representation (Putzar et al., 2007; Pasqualotto and Proulx, 2012; Gori et al., 2014).

In sum, our study offers a *potential* new avenue for reducing the number of training sessions necessary for

using The vOICe in real-world settings and it could help a conspicuous portion of visually impaired individuals to improve their mobility and social interactions (Dundon et al., 2015).

AUTHOR CONTRIBUTIONS

AP followed every phase of the work; TE contributed to collect data and to discuss/interpret the results.

REFERENCES

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, 687–689. doi: 10.1038/nn1912

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–266; discussion 266–313. doi: 10.1017/S0140525X10000853

Arditi, A., Holtzman, J. D., and Kosslyn, S. M. (1988). Mental imagery and sensory experience in congenital blindness. *Neuropsychologia* 26, 1–12. doi: 10.1016/0028-3932(88)90026-7

Auvray, M., Hanneton, S., Lenay, C., and O'Regan, K. (2005). There is something out there: distal attribution in sensory substitution, twenty years later. *J. Integr. Neurosci.* 4, 505–521. doi: 10.1142/s0219635205001002

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. doi: 10.1068/p5631

Avenanti, A., Annela, L., and Serino, A. (2012). Suppression of premotor cortex disrupts motor coding of peripersonal space. *Neuroimage* 63, 281–288. doi: 10.1016/j.neuroimage.2012.06.063

Avraamides, M. N., and Kelly, J. W. (2010). Multiple systems of spatial memory: evidence from described scenes. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 635–645. doi: 10.1037/a0017040

Avraamides, M. N., Loomis, J. M., Klatzky, R. L., and Golledge, R. G. (2004). Functional equivalence of spatial representations derived from vision and language: evidence from allocentric judgments. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 801–814. doi: 10.1037/0278-7393.30.4.804

Bach-y-Rita, P. (1972). *Brain Mechanisms in Sensory Substitution.* New York: Academic Press.

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. doi: 10.1038/221963a0

Bach-y-Rita, P., and Kercel, S. W. (2003). Sensory substitution and the human-machine interface. *Trends Cogn. Sci.* 7, 541–546. doi: 10.1016/j.tics.2003.10.013

Ballesteros, S., Millar, S., and Reales, J. M. (1998). Symmetry in haptic and in visual shape perception. *Percept. Psychophys.* 60, 389–404. doi: 10.3758/bf03206862

Beer, A. L., Plank, T., and Greenlee, M. W. (2011). Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. *Exp. Brain Res.* 213, 299–308. doi: 10.1007/s00221-011-2715-y

Brown, D. J., Macpherson, T., and Ward, J. (2011). Seeing with sound? Exploring different characteristics of a visual-to-auditory sensory substitution device. *Perception* 40, 1120–1135. doi: 10.1068/p6952

Brown, D. J., Simpson, A. J. R., and Proulx, M. J. (2015). Auditory scene analysis and sonified visual images. Does consonance negatively impact on object formation when using complex sonified stimuli? *Front. Psychol.* 6:1522. doi: 10.3389/fpsyg.2015.01522

Burgess, N., Spiers, H. J., and Paleologou, E. (2004). Orientational manoeuvres in the dark: dissociating allocentric and egocentric influences on spatial memory. *Cognition* 94, 149–166. doi: 10.1016/j.cognition.2004.01.001

Chebat, D. R., Schneider, F. C., Kupers, R., and Ptito, M. (2011). Navigation with a sensory substitution device in congenitally blind individuals. *Neuroreport* 22, 342–347. doi: 10.1097/WNR.0b013e3283462def

Cramer, D. (1998). *Fundamental Statistics for Social Research: Step-by-step Calculations and Computer Techniques Using SPSS for Windows.* London: Routledge.

De Volder, A. G., Toyama, H., Kimura, Y., Kiyosawa, M., Nakano, H., Vanlierde, A., et al. (2001). Auditory triggered mental imagery of shape involves visual association areas in early blind humans. *Neuroimage* 14, 129–139. doi: 10.1006/nimg.2001.0782

Diwadkar, V. A., and McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychol. Sci.* 8, 302–307. doi: 10.1111/j.1467-9280.1997.tb00442.x

Dundon, N. M., Bertini, C., Làdavas, E., Sabel, B. A., and Gall, C. (2015). Visual rehabilitation: visual scanning, multisensory stimulation and vision restoration trainings. *Front. Behav. Neurosci.* 9:192. doi: 10.3389/fnbeh.2015.00192

Farrell, M. J., and Robertson, I. H. (2000). The automatic updating of egocentric spatial relationships and its impairment due to right posterior cortical lesions. *Neuropsychologia* 38, 585–595. doi: 10.1016/s0028-3932(99)00123-2

Farrell, M. J., and Thomson, J. A. (1998). Automatic spatial updating during locomotion without vision. *Q. J. Exp. Psychol. A* 51, 637–654. doi: 10.1080/713755776

Ganis, G., Thompson, W. L., and Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Brain Res. Cogn. Brain Res.* 20, 226–241. doi: 10.1016/j.cogbrainres.2004.02.012

Gaunet, F., Vidal, M., Kemeny, A., and Berthoz, A. (2001). Active, passive and snapshot exploration in a virtual environment: influence on scene memory, reorientation and path memory. *Brain Res. Cogn. Brain Res.* 11, 409–420. doi: 10.1016/s0926-6410(01)00013-1

Ghazanfar, A. A., and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* 10, 278–285. doi: 10.1016/j.tics.2006.04.008

Giudice, N. A., Betty, M. R., and Loomis, J. M. (2011). Functional equivalence of spatial images from touch and vision: evidence from spatial updating in blind and sighted individuals. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 621–634. doi: 10.1037/a0022331

Gori, M., Sandini, G., Martinoli, C., and Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Curr. Biol.* 20, 223–225. doi: 10.1016/j.cub.2009.11.069

Gori, M., Sandini, G., Martinoli, C., and Burr, D. C. (2014). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain* 137, 288–293. doi: 10.1093/brain/awt311

Haigh, A., Brown, D. J., Meijer, P., and Proulx, M. J. (2013). How well do you see what you hear? The acuity of visual-to-auditory sensory substitution. *Front. Psychol.* 4:330. doi: 10.3389/fpsyg.2013.00330

Hartley, T., Lever, C., Burgess, N., and O'Keefe, J. (2014). Space in the brain: how the hippocampal formation supports spatial cognition. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20120510. doi: 10.1098/rstb.2012.0510

Ho, C., and Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *J. Exp. Psychol. Appl.* 11, 157–174. doi: 10.1037/1076-898x.11.3.157

Intraub, H., Morelli, F., and Gagnier, K. M. (2015). Visual, haptic and bimodal scene perception: evidence for a unitary representation. *Cognition* 138, 132–147. doi: 10.1016/j.cognition.2015.01.010

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., and Hudspeth, A. J. (2012). *Principles of Neural Science.* New York: McGraw-Hill.

Kappers, A. M., and Koenderink, J. J. (1999). Haptic perception of spatial relations. *Perception* 28, 781–795. doi: 10.1068/p2930

Kim, J. K., and Zatorre, R. J. (2008). Generalized learning of visual-to-auditory substitution in sighted individuals. *Brain Res.* 1242, 263–275. doi: 10.1016/j.brainres.2008.06.038

King, A. J. (2009). Visual influences on auditory spatial learning. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 331–339. doi: 10.1098/rstb.2008.0230

Klatzky, R. L., Lippa, Y., Loomis, J. M., and Golledge, R. G. (2003). Encoding, learning and spatial updating of multiple object locations specified by 3-D sound, spatial language and vision. *Exp. Brain Res.* 149, 48–61. doi: 10.1007/s00221-002-1334-z

Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., and Golledge, R. G. (1998). Spatial updating of self-position and orientation during real, imagined

and virtual locomotion. *Psychol. Sci.* 9, 293–298. doi: 10.1111/1467-9280.00058

Knudsen, E. I., and Brainard, M. S. (1995). Creating a unified representation of visual and auditory space in the brain. *Ann. Rev. Neurosci.* 18, 19–43. doi: 10.1146/annurev.neuro.18.1.19

Lacey, S., Campbell, C., and Sathian, K. (2007). Vision and touch: multiple or multisensory representations of objects? *Perception* 36, 1513–1522. doi: 10.1068/p5850

Lenay, C., Gapenne, O., Hanneton, S., Genouëlle, C., and Marque, C. (2003). "Sensory substitution: limits and perspectives," in *Touching for Knowing*, eds Y. Hatwell, A. Streri, and E. Gentaz (Amsterdam: John Benjamins), 275–292.

Lomber, S. G., and Malhotra, S. (2008). Double dissociation of "what" and "where" processing in auditory cortex. *Nat. Neurosci.* 11, 609–616. doi: 10.1038/nn.2108

Loomis, J. M. (2010). "Sensory substitution for orientation and mobility: what progress are we making?," in *Foundations of Orientation and Mobility*, eds W. R. Wiener, R. L. Welsh, and B. B. Blasch (New York, U S A: AFB Press), 3–44.

Loomis, J. M., Klatzky, R. L., and Giudice, N. A. (2013). "Representing 3D space in working memory: spatial images from vision, hearing, touch and language," in *Multisensory Imagery*, eds S. Lacey and R. Lawson (New York, U S A: Springer), 131–155.

Loomis, J. M., Lippa, Y., Klatzky, R. L., and Golledge, R. G. (2002). Spatial updating of locations specified by 3-D sound and spatial language. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 335–345. doi: 10.1037/0278-7393.28.2.335

Maidenbaum, S., Abboud, S., and Amedi, A. (2014). Sensory substitution: closing the gap between basic research and widespread practical visual rehabilitation. *Neurosci. Biobehav. Rev.* 41, 3–15. doi: 10.1016/j.neubiorev.2013.11.007

Makin, T. R., Holmes, N. P., and Zohary, E. (2007). Is that near my hand? Multisensory representation of peripersonal space in human intraparietal sulcus. *J. Neurosci.* 27, 731–740. doi: 10.1523/JNEUROSCI.3653-06.2007

McNamara, T. P. (2003). "How are locations of objects in the environment represented in memory?," in *Spatial Cognition, III: Routes and Navigation, Human Memory and Learning, Spatial Representation and Spatial Reasoning*, eds C. Freska, W. Brauer, C. Habel and K. Wender (Berlin, Germany: Springer), 174–191.

Meijer, P. B. (1992). An experimental system for auditory image representations. *IEEE Trans. Biomed. Eng.* 39, 112–121. doi: 10.1109/10.121642

Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision: two central pathways. *Trends Neurosci.* 6, 414–417. doi: 10.1016/0166-2236(83)90190-x

Morris, A. P., Chambers, C. D., and Mattingley, J. B. (2007). Parietal stimulation destabilizes spatial updating across saccadic eye movements. *Proc. Nat. Acad. Sci. U S A* 104, 9069–9074. doi: 10.1073/pnas.0610508104

Mou, W., Hayward, W. G., Zhao, M., Zhou, G., and Owen, C. B. (2006). Spatial updating during locomotion does not eliminate viewpoint-dependent visual object processing. *J. Vis.* 6:316. doi: 10.1167/6.6.316

Mou, W., and McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 162–170. doi: 10.1037/0278-7393.28.1.162

Mou, W., McNamara, T. P., Valiquette, C. M., and Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 142–157. doi: 10.1037/0278-7393.30.1.142

Newell, F. N., Woods, A. T., Mernagh, M., and Bülthoff, H. H. (2005). Visual, haptic and crossmodal recognition of scenes. *Exp. Brain Res.* 161, 233–242. doi: 10.1007/s00221-004-2067-y

Papadopoulos, K., Koustriava, E., and Kartasidou, L. (2012). Spatial coding of individuals with visual impairments. *J. Spec. Educ.* 46, 180–190. doi: 10.1177/0022466910383016

Pascual-Leone, A., and Hamilton, R. (2001). The metamodal organization of the brain. *Progr. Brain Res.* 134, 427–445. doi: 10.1016/s0079-6123(01)34028-1

Pasqualotto, A., Dumitru, M. L., and Myachykov, A. (2015). Editorial: multisensory integration: brain, body and world. *Front. Psychol.* 6:2046. doi: 10.3389/fpsyg.2015.02046

Pasqualotto, A., Finucane, C. M., and Newell, F. N. (2005). Visual and haptic representations of scenes are updated with observer movement. *Exp. Brain Res.* 166, 481–488. doi: 10.1007/s00221-005-2388-5

Pasqualotto, A., Finucane, C. M., and Newell, F. N. (2013a). Ambient visual information confers a context-specific, long-term benefit on memory for haptic scenes. *Cognition* 128, 363–379. doi: 10.1016/j.cognition.2013.04.011

Pasqualotto, A., Spiller, M. J., Jansari, A. S., and Proulx, M. J. (2013b). Visual experience facilitates allocentric spatial representation. *Behav. Brain Res.* 236, 175–179. doi: 10.1016/j.bbr.2012.08.042

Pasqualotto, A., and Proulx, M. J. (2012). The role of visual experience for the neural basis of spatial cognition. *Neurosci. Biobehav. Rev.* 36, 1179–1187. doi: 10.1016/j.neubiorev.2012.01.008

Pasqualotto, A., and Proulx, M. J. (2013). The study of blindness and technology can reveal the mechanisms of three-dimensional navigation. *Behav. Brain Sci.* 36, 559–560; discussion 571–587. doi: 10.1017/s0140525x13000496

Proulx, M. J., Brown, D. J., Pasqualotto, A., and Meijer, P. (2014a). Multisensory perceptual learning and sensory substitution. *Neurosci. Biobehav. Rev.* 41, 16–25. doi: 10.1016/j.neubiorev.2012.11.017

Proulx, M. J., Ptito, M., and Amedi, A. (2014b). Multisensory integration, sensory substitution and visual rehabilitation. *Neurosci. Biobehav. Rev.* 41, 1–2. doi: 10.1016/j.neubiorev.2014.03.004

Proulx, M. J., Stoerig, P., Ludowig, E., and Knoll, I. (2008). Seeing 'where'through the ears: effects of learning-by-doing and long-term sensory deprivation on localization based on image-to-sound substitution. *Plos One* 3:e1840. doi: 10.1371/journal.pone.0001840

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. doi: 10.1093/brain/awh380

Putzar, L., Goerendt, I., Lange, K., Rösler, F., and Röder, B. (2007). Early visual deprivation impairs multisensory interactions in humans. *Nat. Neurosci.* 10, 1243–1245. doi: 10.1038/nn1978

Rauschecker, J. P. (1995). Compensatory plasticity and sensory substitution in the cerebral cortex. *Trends Neurosci.* 18, 36–43. doi: 10.1016/0166-2236(95)93948-w

Sadato, N., Nakashita, S., and Saito, D. N. (2007). Pathways of tactile-visual crossmodal interaction for perception. *Behav. Brain Sci.* 30, 218–219. doi: 10.1017/s0140525x07001586

Sakata, H., and Kusunoki, M. (1992). Organization of space perception: neural representation of three-dimensional space in the posterior parietal cortex. *Curr. Opin. Neurobiol.* 2, 170–174. doi: 10.1016/0960-9822(92)90356-f

Sampaio, E., Maris, S., and Bach-y-Rita, P. (2001). Brain plasticity: 'Visual' acuity of blind persons via the tongue. *Brain Res.* 908, 204–207. doi: 10.1016/s0006-8993(01)02667-1

Schifferstein, H. N. J., Smeets, M. A. M., and Postma, A. (2010). Comparing location memory for 4 sensory modalities. *Chem. Senses* 35, 135–145. doi: 10.1093/chemse/bjp090

Shelton, A. L., and McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cogn. Psychol.* 43, 274–310. doi: 10.1006/cogp.2001.0758

Simons, D. J., and Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychol. Sci.* 9, 315–320. doi: 10.1111/1467-9280.00062

Stiles, N. R., Zheng, Y., and Shimojo, S. (2015). Length and orientation constancy learning in 2-dimensions with auditory sensory substitution: the importance of self-initiated movement. *Front. Psychol.* 6:842. doi: 10.3389/fpsyg.2015.00842

Striem-Amit, E., Guendelman, M., and Amedi, A. (2012). 'Visual' acuity of the congenitally blind using visual-to-auditory sensory substitution. *Plos One* 7:e33136. doi: 10.1371/journal.pone.0033136

Thibault, G., Pasqualotto, A., Vidal, M., Droulez, J., and Berthoz, A. (2013). How does horizontal and vertical navigation influence spatial memory of multifloored environments? *Atten. Percept. Psychophys.* 75, 10–15. doi: 10.3758/s13414-012-0405-x

Tokumaru, O., Mizumoto, C., Takada, Y., and Ashida, H. (2003). EEG activity of aviators during imagery flight training. *Clin. Neurophysiol.* 114, 1926–1935. doi: 10.1016/s1388-2457(03)00172-x

Tyler, M., Danilov, Y., and Bach-y-Rita, P. (2003). Closing an open-loop control system: vestibular substitution through the tongue. *J. Integr. Neurosci.* 2, 159–164. doi: 10.1142/s0219635203000263

Waller, D., Montello, D. R., Richardson, A. E., and Hegarty, M. (2002). Orientation specificity and spatial updating of memories for layouts. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 1051–1063. doi: 10.1037/0278-7393.28.6.1051

White, B. W., Saunders, F. A., Scadden, L., Bach-Y-Rita, P., and Collins, C. C. (1970). Seeing with the skin. *Percept. Psychophys.* 7, 23–27. doi: 10.3758/BF03210126

Wolbers, T., and Büchel, C. (2005). Dissociable retrosplenial and hippocampal contributions to successful formation of survey representations. *J. Neurosci.* 25, 3333–3340. doi: 10.1523/JNEUROSCI.4705-04.2005

Woods, A. T., and Newell, F. N. (2004). Visual, haptic and cross-modal recognition of objects and scenes. *J. Physiol. Paris* 98, 147–159. doi: 10.1016/j.jphysparis.2004.03.006

Yamamoto, N., and Shelton, A. L. (2009). Orientation dependence of spatial memory acquired from auditory experience. *Psychon. Bull. Rev.* 16, 301–305. doi: 10.3758/PBR.16.2.301

Zhao, M., Zhou, G., Mou, W., Hayward, W. G., and Owen, C. B. (2007). Spatial updating during locomotion does not eliminate viewpoint-dependent visual object processing. *Vis. Cogn.* 15, 402–419. doi: 10.1080/13506280600783658

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.