

# Ears on the Hand: Reaching Three-Dimensional Targets With an Audio-Motor Device

Sylvain Hanneton<sup>1</sup>, Thomas Hoellinger<sup>2</sup>, Vincent Forma<sup>3</sup>,  
Agnes Roby-Brami<sup>4,5</sup> and Malika Auvray<sup>4,\*</sup>

<sup>1</sup> Institut des Sciences du Sport-Santé EA3625, Université Paris Descartes, Paris, France

<sup>2</sup> Laboratoire de Neurophysiologie et Biomécanique du mouvement, Faculté des Sciences de la motricité, Université Libre de Bruxelles, Brussel, Belgium

<sup>3</sup> Laboratoire Psychologie de la Perception, CNRS UMR 8242, Université Paris Descartes, Paris, France

<sup>4</sup> Institut des Systèmes Intelligents et de Robotique, ISIR, CNRS UMR 7222, Sorbonne Université, Paris, France

<sup>5</sup> Institut des Systèmes Intelligents et de Robotique, Equipe Agathe, INSERM U 1150, Paris, France

Received 11 February 2019; accepted 18 September 2109

---

## Abstract

Understanding the processes underlying sensorimotor coupling with the environment is crucial for sensorimotor rehabilitation and sensory substitution. In doing so, devices which provide novel sensory feedback consequent to body movement may be optimized in order to enhance motor performance for particular tasks. The aim of the study reported here was to investigate audio-motor coupling when the auditory experience is linked to movements of the head or the hands. The participants had to localize and reach a virtual source with the dominant hand in response to sounds. An electromagnetic system recorded the position and orientation of the participants' head and hands. This system was connected to a 3D audio system that provided binaural auditory feedback on the position of the virtual listener located on the participants' body. The listener's position was computed either from the hands or from the head. For the hand condition, the virtual listener was placed on the dominant hand (the one used to reach the target) in Experiment 1 and on the non-dominant hand, which was constrained in order to have similar amplitude and degrees of freedom as that of the head, in Experiment 2. The results revealed that, in the two experiments, the participants were able to localize a source within the 3D auditory environment. Performance varied as a function of the effector's degrees of freedom and the spatial coincidence between sensor and effector. The results also allowed characterizing the kinematics of the hand and head and how they change with audio-motor coupling condition and practice.

---

\* To whom correspondence should be addressed. E-mail: [auvray@isir.upmc.fr](mailto:auvray@isir.upmc.fr)

**Keywords**

Sensory substitution, neural plasticity, perceptual adaptation, rehabilitation

**1. Introduction***1.1. Sensory Substitution*

In the field of perceptual rehabilitation, sensory substitution devices were developed with the aim of compensating one or several functions of a deficient sensory modality. To do so, these devices convert stimuli that are normally accessed through this deficient sensory modality into stimuli accessible by another sensory modality (e.g., Auvray and Myin, 2008; Gallay *et al.*, 2012, for reviews). For instance, these devices can convert visual information into tactile stimuli (e.g., Bach-y-Rita *et al.*, 1969) or sounds (e.g., Hanneton *et al.*, 2010; Meijer, 1992). Studies conducted on sensory substitution revealed an important plasticity of the central nervous system. Behavioural studies revealed that sensory substitution devices allow improving spatial processing in the absence of vision. In particular, users of these systems become able to perform localisation tasks (e.g., Levy-Tzedek *et al.*, 2012; Proulx *et al.*, 2008), recognition tasks (Arno *et al.*, 2001; Auvray *et al.*, 2007a, b; Pollok *et al.*, 2005; Sampaio *et al.*, 2001), navigation tasks (Chebat *et al.*, 2011, 2015) and, to some extent, consolidate a perceptual space that possesses visual characteristics (e.g., Renier *et al.*, 2005). However, the degree to which the use of a sensory substitution device involves visual processes varies from one person to another (Arnold *et al.*, 2017). In parallel, brain-imaging studies have shown that the use and practice of visual-to-tactile (Ptito *et al.*, 2005) and visual-to-auditory (e.g., Striem-Amit and Amedi, 2014) devices result in increased activation in blind people's visual cortex (for reviews, see Proulx *et al.*, 2014; Ptito *et al.*, 2018).

*1.2. Sensory Substitution, Learning, and Body Kinematics*

Effective use of sensory substitution devices is largely dependent upon the users' ability to master their different features and the amount of training required to do so. In most cases, users are able to approximate objects' location and shape in a short amount of time (i.e., one or two hours with the Tongue Display Unit, Kaczmarek and Haase, 2003). However, to gain precision in these tasks and to perform complex ones, more intensive training is required. The length of the training is estimated at around 10 to 15 hours with visual-to-auditory devices (Auvray *et al.*, 2007) and around eight hours with visual-to-tactile devices (Kaczmarek and Haase, 2003). It is estimated that further time still would be required to achieve a high level of expertise,

for instance to be able to recognize facial expressions with visual-to-auditory devices (Striem-Amit *et al.*, 2012).

Another crucial characteristic of learning is the ability to generalize, that is, the ability to extend the acquired perceptual abilities to both new stimuli and new perceptual conditions. With tactile devices, it has been shown that learning transfers across body surfaces, both for surfaces represented in adjacent and non-adjacent areas of the somatosensory cortex (Arnold and Auvray, 2014). This result provides behavioural support for the claim that training in sensory substitution results in a relative independence from the stimulated body surface. In addition, transfer of learning to novel stimuli has been documented for both tactile (Arnold and Auvray, 2018; Kim and Zatorre, 2010) and auditory devices (Kim and Zatorre, 2008). This ability to generalize to new symbols depends on the number of stimuli to learn and on feature variability (Arnold and Auvray, 2018). These results are in line with the reverse hierarchy theory (RHT) of perceptual learning (Ahissar and Hochstein, 2004), according to which the degree of specificity depends on the difficulty of the perceptual task and on the level of cortical processing that is required to perform the task.

Finally, performance and experience with sensory substitution devices depend on each person's abilities (see Arnold *et al.*, 2017, for a review). For instance, Brown *et al.*'s (2011) study showed that people with good musical abilities learned a visual-to-auditory device faster than people with low musical abilities, although they did not reach better abilities with the device. Haigh *et al.*'s (2013) study found that musicians had a better "visual" acuity (tested with Snellen E) when using the visual-to-auditory conversion system The vOICE than non-musicians. Note that regarding visual abilities, Brown *et al.* (2011) found no influence of individual visual imagery abilities (measured with VVIQ, Marks, 1973) on the learning of the vOICE. On the other hand, in a study by Hanne-ton *et al.* (2015), performance reached by the participants on a task based on the auditory conversion of a visual video game was found to be correlated to the scores the participants obtained in a visual mental rotation test (Vandenberg, 1971). It thus seems that performance with sensory substitution devices is related to only some visuo-spatial abilities. Overall, these results suggest that perceptual strategies will differ as a function of people's specificities and cognitive style which, in some tasks, affects people's performance.

The studies described above showed people's ability to learn how to use sensory substitution devices. These studies focused on the sensory aspects, leaving the motor changes involved in learning undocumented. With other devices, learning the new sensory consequences of body movements has been shown to increase motor performance and to induce an evolution of body movement characteristics. In particular, as adaptation progresses, movements become less segmented, smoother, more rapid and fluid, with an increase in the amplitude of joint motion (e.g., Novak *et al.*, 2003; Rand *et al.*, 2004). Our

study aimed at investigating such motor consequences of perceptual learning in the case of sensory substitution.

### *1.3. Aim of the Study*

The first aim of the study reported here was to investigate the changes in kinematic parameters while learning a novel auditory device. To do so, the participants had to learn to intercept a static auditory target within a 3D immersive auditory environment. Motion capture sensors were placed on the participants' head and hand, and their kinematics were analysed while adapting to this new system. In addition to success rate and reaching duration, the kinematics of hand and head movements, and in particular the number of velocity peaks, the direct distance followed by the hand and head during reaching, the cumulative length of the trajectories, and the cumulative rotations were analysed.

The second aim of the study was to investigate if performance with the device is influenced by the location of the sensors. In particular, one hypothesis raised by sensorimotor theories of perception is that the use of a head-mounted camera shares more sensorimotor resemblance with vision than the use of a hand-held camera (e.g., O'Regan, 2011; O'Regan and Noë, 2001). As a consequence, according to this theory, the use of a head-mounted camera might favour an association between perception with the device and the visual sensory modality as compared with the situations in which the camera is hand-held. Thus, two modes were investigated in our experiment, the head mode and the hand mode. In the head mode, the head sensor was used to drive the position and orientation of the virtual listener in the 3D auditory display. In the hand mode, the position and orientation of the virtual listener was driven by the sensor fixed on the hand. Consequently, in this condition, the target's sound is perceived through "ears" on the participant's hand.

Thirdly, in order to further investigate the influence of body kinematics on the use of the device, our experiments studied the extent to which the differences in performance obtained with the hand and head is attributable to differences in amplitude of movement and degrees of freedom. Thus, in the first experiment, hand movements were not constrained, whereas in the second experiment, the movements of the non-dominant hand wearing the listener were constrained in order to have similar amplitude and degrees of freedom as that of the head.

## **2. Experiment 1**

### *2.1. Materials and Methods*

#### *2.1.1. Participants*

12 participants (4 females, 2 left-handed, age range: 25–50 years) took part in this experiment which lasted approximately 40 min. All were students or

faculty members at Paris Descartes University. The participants were aged between 18 and 50 years, in order to avoid any age-related hearing loss. All the participants self-reported neither hearing impairment nor neurological disorder. This experiment was conducted in accordance with the ethical standards laid down in the 1991 Declaration of Helsinki.

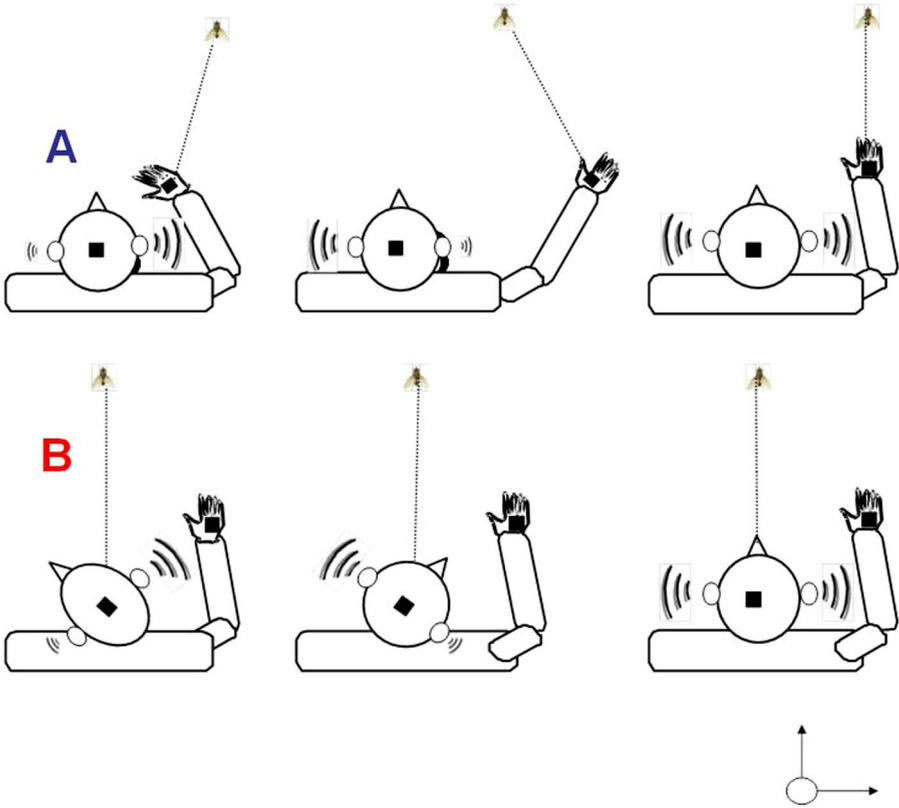
### *2.1.2. Experimental Set-up*

An electromagnetic device (Fastrak Polhemus) was connected to the 3D audio system OpenAL that provides auditory feedback on movements. The Polhemus system acquires 6D motion data with a 30 Hz sample frequency. This electromagnetic system gives real time position (X, Y, Z) and orientation (Euler angles: azimuth, elevation, and roll) of two sensors, allowing a distinction between changes in position and orientation. The body-auditory coupling system is thus made of a listener (sensor) placed on the participants' body and a virtual audio source. The perceived sounds depend on the position and orientation of the listener relative to the position of the source (see Fig. 1). More precisely, a "buzzing" sound similar to a fly with significant harmonics between 100 and 2000 Hz varied in intensity and frequency as a function of the position and orientation of the listener (see below for the modelling and equations used).

In the "head mode", the listener was attached to the head and the perceived sound was sensitive to the position and orientation of the head with respect to the sound source. It corresponded to a quite natural situation since the positions of the "virtual ears" were located close to the real ears. In the "hand mode", the listener was attached to the hand used to reach the fly and the perceived sound was sensitive to the position and orientation of the hand relative to the presented sound source (see Fig. 1).

### *2.1.3. Sound Spatialisation Modelling*

The binaural sound generated by the audio device (SoundMax integrated digital audio by Analog Devices Inc.) was computed from the relative position and orientation between the listener and the virtual source (see Fig. 1). The 3D audio system was based on the library OpenAL (see Note 1), used in video games, which allows 3D sound spatialisation. OpenAL uses both interaural intensity differences (IIDs) and interaural time differences (ITDs), which are combined to code for the target's distance and direction. As a sound source does not have the same distance between the two ears, the sound arrives faster to the closest ear (ITDs) and with a higher intensity (IIDs). Regarding IIDs, according to the specifications of OpenAL library, the intensity decreases and the sound is slightly muffled as the listener moves away from the source. This models reality in which air absorbs high frequency waves more rapidly than low frequency ones. Within 5 cm from the source, the sound remains at a



**Figure 1.** Schematic representation of the experimental set-up. The auditory target is represented by a fly. For each mode, the sensor (or listener) is represented by a black square placed either on the dominant hand (A) or on the head (B). The schema illustrates how the participants' hand or head orientation relative to the sound source is used to produce spatialised auditory feedback.

constant maximum volume. Beyond a maximal distance of 40 cm, the amplitude is held constant. Between these two limits, the intensity of the sound is computed as a function of distance, for each hear, according to the following equation:  $G(d) = D_{ref}/(D_{ref} + R * (d - D_{ref}))$ , where  $d$  is the distance to the source,  $G$  is the gain applied to the sound,  $R (= 1.3)$  is the Roll-off factor, and  $D_{ref} (= 5 \text{ cm})$  is a reference distance corresponding to the distance that gives a unitary gain.

**2.1.4. Procedure**

For each of the two modes (head and hand), the participants were instructed to catch the fly, i.e., to reach the static virtual sound source, with their dominant hand. The participants sat in front of a table. They were blindfolded, fitted with headphones, and two sensors were fixed with specific double-sided tape

on the back of the hand and on the headphones. A geometrical transformation was applied in order to place the virtual listener either on the centre of the hand or on the centre of the head, midway between the ears. At the start of each trial, the dominant hand, used to reach the source, was placed on the sternum near the xyphoid process. The participants were instructed to start to move as soon as they heard the sound. The trial was considered as being successful (i.e., the “fly” was caught) when the centre of the hand approached within a distance smaller than 5 cm of the static sound source. In this case, the recording stopped and a synthetic applause was played. The participants were able to give their responses up to 31 s after the end of the stimulus.

The virtual source could be presented at 9 possible locations within the participants’ workspace: 1 in the middle and 8 in the corners of a virtual parallelepiped rectangle (20 cm on the right or the left of the participants’ midline, at 5 or 25 cm from of their sternum and at 15 or 35 cm above the table). The order of presentation of the 9 targets was randomized within blocks. Blocks of 9 trials were repeated three times for each of the two modes, with the participants completing 54 trials in total. Half of the participants began with the hand mode while the other half began with the head mode.

#### 2.1.5. Statistical Analyses

The computed dependent variables were the success rate (percentage of catches before cut-off) and reaching duration (time to reach the target). The kinematics of hand and head movements during successful trials were characterized by the number of velocity peaks, the direct distance followed by the hand and head during reaching (3D distance between their starting and final positions), each sensor’s cumulative length of trajectories and cumulative rotations. The cumulative length of the trajectories corresponds to the sum of all the elementary 3D displacements performed during each 33 ms time sample. The cumulative rotations were obtained after calculating the 3D rotation matrix at each time sample, then the change in 3D orientation during one time sample. The rotation angle performed during each time sample was computed using the axis-angle formalism. Then, the obtained elementary rotations were summed during the whole trial to obtain the cumulative rotations of the hand and head.

The distribution of the variables was verified with Mauchley’s Sphericity test. Repeated measures ANOVAs were conducted to analyse the dependent variables with the factors Session (2 levels), Mode (2 levels: hand and head), and Block (3 levels). The results were averaged across targets. *Post Hoc* comparisons were performed with Bonferroni tests. A Multivariate approach of ANOVA was used when the conditions for ANOVA were not fully met, completed with Friedman non-parametric analysis, when relevant. Additional statistics were performed using Chi-squared test.

## 2.2. Results

### 2.2.1. Success Rate

The success rate appeared to be variable across participants and conditions. The mean success rate in the three blocks ranged between 51.8 and 100% in the hand mode and between 26 and 100% in the head mode. The success rate did not vary with the Order of sessions and there was no interaction with Mode (Pearson's Chi-squared test,  $p = 0.18$  and  $= 0.38$ , respectively). Given this, results were pooled across sessions for the subsequent analyses.

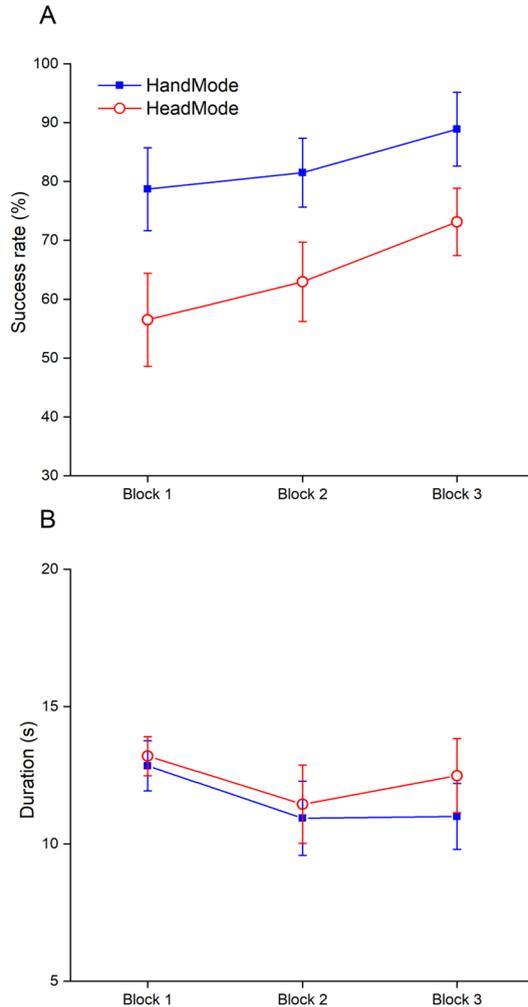
An ANOVA conducted with the factors Mode and Block showed a significant effect of Mode ( $F[1, 11] = 15.03$ ,  $p = 0.002$ ). The success rate was higher in the hand mode ( $75.93 \pm 4.79\%$ , mean  $\pm$  sem) than in the head mode ( $71.3 \pm 7.58\%$ ) (Fig. 2A). The sphericity condition was not fully met for the Block factor (Mauchly's test  $p = 0.044$ ). Non parametric statistics (Friedman test) showed a significant effect of Block in the head mode ( $p = 0.011$ ) but not in the hand mode ( $p = 0.30$ ).

### 2.2.2. Trial Duration

An ANOVA conducted with the Factors Mode and Block on trial duration did not show any significant main effect, nor interaction between Mode and Block (all  $ps > 0.05$ ). Numerically, trial duration was  $11.59 \pm 1.03$  s in the hand mode and  $12.37 \pm 1.02$  s in the head mode (see Fig. 2B).

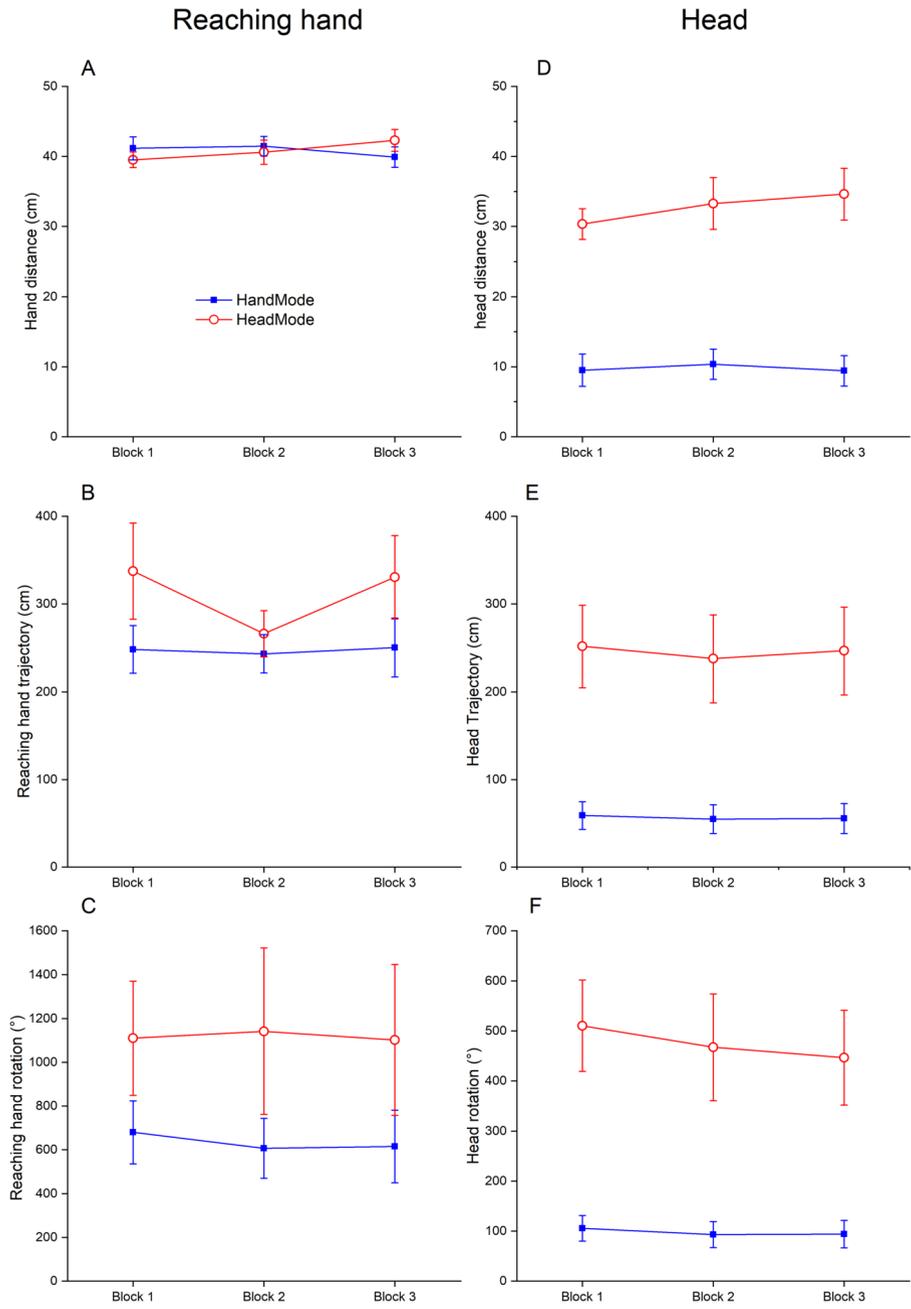
### 2.2.3. Hand and Head Kinematics

An ANOVA conducted with the Factors Mode and Block on the direct distance followed by the hand during reaching did not show any effect of Mode nor Block ( $Fs < 1$ ), confirming that movement execution during successful trials was reproducible across conditions (see Fig. 3A). The trajectories followed by the head and hand did not follow a single path, but several of them, suggesting a behaviour of exploration. As a consequence, cumulative trajectories and cumulative rotations of the head and hand were used as dependent variables. Figure 3 shows the kinematic variables of the reaching hand: cumulative trajectories (Fig. 3B) and cumulative rotations (Fig. 3C). The observation of the graphs suggests that the participants followed longer exploratory movements of the reaching right hand, with more cumulative rotation in the head mode than in the hand mode. Interestingly, the variability was more important in the head mode, and in particular regarding the cumulative hand rotations. ANOVAs were conducted with the factors Mode and Block on the dependent variables describing reaching hand trajectories: cumulative trajectories, number of velocity peaks, and cumulative rotations of the reaching hand trajectory. The results for cumulative trajectories failed to show any significant effect of Mode ( $F[1, 11] = 4.28$ ,  $p = 0.062$ ) or Block ( $F[2, 22] = 1.12$ ,  $p = 0.34$ ). Similarly, the results on number of velocity



**Figure 2.** Success rate and trial duration in Experiment 1. The graphs show the evolution of the success rate (A) and trial duration (B) during the three successive blocks for the two modes: hand mode (in blue) and head mode (in red). The symbols and vertical lines represent the mean and standard error to the mean.

peaks failed to show any significant main effect of Mode ( $F[1, 11] = 0.45$ ,  $p = 0.51$ ) or Block ( $F[2, 22] = 2.35$ ,  $p = 0.11$ ). The results on cumulative rotations of the reaching hand trajectory showed a significant effect of Mode ( $F[1, 11] = 5.8$ ,  $p = 0.034$ ), no effect of Block ( $F[2, 22] = 0.6$ ,  $p = 0.55$ ), and no interaction between Mode and Block ( $F[2, 22] = 0.71$ ,  $p = 0.50$ ). There were more cumulative rotations of the reaching hand trajectory in the head mode than in the hand mode (see Fig. 3C).



The direct distance followed by the head and its kinematic variables: cumulative trajectories and cumulative rotations are shown on Fig. 3D–F. The direct distance followed by the head during reaching showed a significant effect of Mode ( $F[1, 11] = 82.06$ ,  $p < 0.0001$ ), no effect of Block ( $F[2, 22] = 1.18$ ,  $p = 0.32$ ), and no interaction between these two factors ( $F[2, 22] = 1.86$ ,  $p = 0.17$ ). The distance was  $9.76 \pm 2.01$  cm in the hand mode and  $32.75 \pm 3.03$  cm in the head mode (see Fig. 3D). The cumulative trajectories and rotations of the head were greater in the head mode ( $171.9 \pm 8.65$  cm and  $319.2 \pm 16.7^\circ$ , respectively) than in the hand mode ( $44 \pm 2.74$  cm and  $74.5 \pm 3.38^\circ$ , respectively), with a higher variability in the head mode (see Fig. 3E, F). The conditions for ANOVA were not fully met (sphericity test  $p = 0.002$ ). Multivariate analysis was used with the factors Block and Mode on the cumulative trajectories. The analysis showed a significant effect of Mode ( $p < 0.0001$ ), no effect of Block ( $p = 0.1$ ), and no interaction between these two factors ( $p = 0.06$ ). Two factors ANOVA for cumulative rotations (Fig. 3F) showed a significant effect of Mode ( $F[1, 11] = 48.36$ ,  $p < 0.0001$ ), no effect of Block ( $F[2, 22] = 0.98$ ,  $p = 0.39$ ), and no interaction between these two factors ( $F[2, 22] = 0.39$ ,  $p = 0.68$ ).

#### 2.2.4. Learning and Adaptation

As was described above, there was no effect of the session and the effect of Block appeared to be complex. Numerically, the success rate increased with Block (Fig. 2A) but the statistics were difficult to interpret, in particular, a borderline sphericity test ( $p = 0.044$ ) weakened the results of the ANOVA. Non parametric Friedman analysis showed an effect of Block in the head mode ( $p = 0.011$ ) but not in the hand mode ( $p = 0.30$ ). It is likely that an important inter-individual variability contributed to this complexity. A possible interpretation is that the lack of significant effect in the hand mode is due to a saturation of the success rate (8 out of 12 participants succeeded in at least 89% of the trials) inducing a ceiling effect, which did not occur in the head mode.

#### 2.2.5. Discussion of Experiment 1

In Experiment 1, the participants were able to use auditory feedback in order to guide reaching movements. In addition, there was a significant improvement in success rate in the head mode and a ceiling effect in the hand mode.

---

**Figure 3.** Kinematics of the reaching hand and of the head in the hand mode (in blue) and head mode (in red). The left part of the figure shows the kinematic variables of the reaching hand: 3D direct distance (A), cumulative trajectories (B), and cumulative rotation (C). The right part of the figure shows the kinematic variables of the head: 3D direct distance (D), cumulative trajectory (E), and cumulative rotation (F). The symbols and vertical lines represent the mean and standard error to the mean.

This result indicates that the participants were able to adapt to a new audio-motor environment and to learn efficient motor strategies. However, most of the kinematic variables did not vary with blocks. A possible explanation is that the participants used quite different individual strategies when completing motor tasks. For example, some of the participants performed relatively direct reaching movements, but most of them used various oscillatory exploratory movements that cannot be easily compared between them.

The second main result from Experiment 1 is that the success rate was higher in the hand mode than in the head mode. In addition, the kinematic analysis of the reaching dominant hand showed that the target was reached after a simpler movement trajectory characterized by less cumulative rotations in the hand mode than in the head mode. This result indicates that when using a sensory substitution device, or any other technology, the idea that sensory organs would benefit from staying at their usual location is not necessarily true, given that here, the most unnatural condition was the most efficient.

It should however be underlined that there were two main differences between the hand mode and the head mode. First, in the hand mode, there was a spatial coincidence between the sensor and effector that was not there in the head mode. Second, the amplitude of possible movements was greater in the hand mode than in the head mode. As a consequence, a second experiment was conducted in order to investigate if these two differences account for the better performance in the hand mode than in the head mode.

### **3. Experiment 2**

The better performance in the hand mode than in the head mode obtained in Experiment 1 can be due to the fact that the head condition involved more complexity and less amplitude of movements. First, in the hand condition, the sensor (the virtual listener) was located on the hand which was also used to catch the fly. Second, head movements have less amplitude than hand movements. To investigate the effect of these differences on participants' performance, in Experiment 2, in the hand condition, the listener was placed on the non-dominant hand while the dominant hand was used to reach the target. In addition, the participants' non-dominant hand displacements were constrained in order to have similar amplitude and degrees of freedom as that of the head.

#### *3.1. Materials and Methods*

##### *3.1.1. Participants*

18 new participants (9 females, 1 left-handed, age range: 20–50 years), recruited at the Paris Descartes University, completed Experiment 2. The participants were aged between 18 and 50 years, in order to avoid any age-related hearing loss. All the participants self-reported neither hearing impairment nor

neurological disorder. This experiment was conducted in accordance with the ethical standards laid down in the 1991 Declaration of Helsinki.

### 3.1.2. Experimental Set-up and Procedure

The same devices and methods as in Experiment 1 were used in Experiment 2, but this time three sensors were used and placed on the two hands and the head, respectively. The reaching movements were again performed by the dominant hand. The listener was defined as the head sensor (head mode, as in Experiment 1) or as the non-dominant hand sensor (non-dominant hand mode). As in Experiment 1, the perceived sounds varied as a function of the sensor's position and orientation with respect to the sound source. In the non-dominant hand mode, the participants were instructed to lean their elbow on a homemade backing fixed on the armrest of the chair, with their forearm elevated so that the hand was at the shoulder's height. This posture aims at limiting the non-dominant hand's degrees of freedom, so that they are similar to those of the head. The conditions of presentation of each auditory target was the same as in Experiment 1: 9 different target positions were presented in each block, in a randomized order. There were 3 successive blocks for each of the two modes with the participants completing 54 trials in total.

### 3.1.3. Performance and Motion Analyses

The same dependent variables as in Experiment 1 were computed: success rate, trial duration, kinematics of head and both hands.

## 3.2. Results

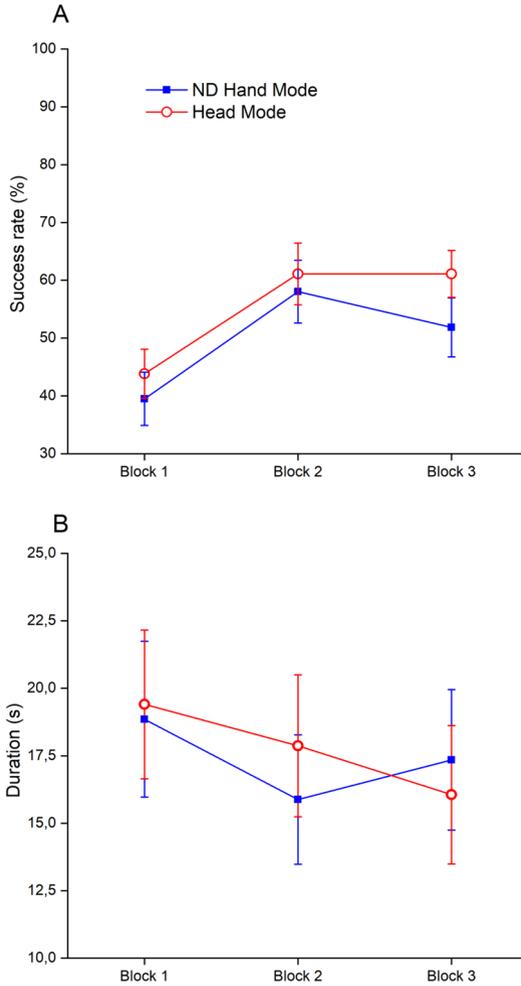
### 3.2.1. Success Rate

The success rate ranged between 14.8 and 77.8% in the hand mode and between 37.0 and 88.9% in the head mode (mean  $\pm$  sem  $55.93 \pm 4.11\%$ , in the hand mode;  $55.35 \pm 3.54\%$  in the head mode). As in Experiment 1, there was no effect of Order of sessions (Pearson's Chi-squared test,  $p = 0.60$ ) and the results were pooled across sessions.

Figure 4A illustrates the evolution of Success rate in the two modes. An ANOVA with the factors Mode and Block showed a significant of Block ( $F[2, 32] = 15.7$ ,  $p < 0.0001$ ), no effect of Mode ( $F[1, 16] = 1.31$ ,  $p = 0.27$ ), and no interaction between these two factors ( $F[2, 32] = 1.04$ ,  $p = 0.36$ ). A Bonferoni *Post-Hoc* test showed that the success rates in Blocks 2 and 3 were significantly higher than in Block 1 ( $p < 0.001$ ).

### 3.2.2. Trial Duration

For Trial duration, the conditions for the ANOVA were not met and a Multivariate analysis failed to show any significant effect of Mode and Block ( $ps > 0.5$ ), nor interaction between these two factors ( $p = 0.19$ ). The numerical trend (illustrated in Fig. 4B) indicated a slight decrease in trial duration with Block, with an increased variability.



**Figure 4.** Success rate and trial duration in Experiment 2. The graphs show the evolution of the success rate (A) and trial duration (B) during the three successive blocks for the two modes: hand mode (in blue) and head mode (in red). The symbols and vertical lines represent the mean and standard error to the mean.

### 3.2.3. Kinematics of the Reaching Right Hand

Figure 5A–C illustrates the kinematic variables of the reaching dominant hand during the task. As in Experiment 1, the direct distance followed by the reaching hand did not vary with the factors Mode and Block, confirming that movement execution during successful trials was reproducible across conditions (see Fig. 5A). The cumulative trajectories and cumulative rotations are shown in Figure 5B, C. An ANOVA conducted with the factors Block and Mode on the cumulative trajectories indicated a borderline effect of Block

( $F[2, 32] = 3.23$ ,  $p = 0.052$ ), no significant effect of Mode ( $F[1, 16] = 3.27$ ,  $p = 0.08$ ), and no interaction between these two factors ( $F[2, 32] = 1.0$ ,  $p = 0.37$ ). An ANOVA conducted with the factors Block and Mode on the cumulative rotations of the reaching hand did not reveal any significant effect of Mode ( $F[1, 16] = 0.43$ ,  $p = 0.8$ ), nor Block ( $F[2, 32] = 1.6$ ,  $p = 0.21$ ), nor interaction ( $F[2, 32] = 0.18$ ,  $p = 0.83$ ). However, the ANOVA conducted on the number of velocity peaks showed a significant effect of Mode ( $F[1, 16] = 8.36$ ,  $p = 0.01$ ), without any effect of Block ( $F[2, 32] = 1.49$ ,  $p = 0.23$ ), nor interaction between these two factors ( $F[2, 32] = 0.46$ ,  $p = 0.63$ ). The number of velocity peaks was higher in the head mode ( $18.2 \pm 1.76$ ) than in the hand mode ( $13.8 \pm 1.44$ ).

#### 3.2.4. Kinematics of the Listener

The listener was the head or the non-dominant hand as a function of the condition. The kinematics of the listener are shown in Fig. 5D-F. A two factors ANOVA on the direct distance followed by the listener (Fig. 5D) showed a significant effect of Mode ( $F[1, 16] = 8.48$ ,  $p = 0.01$ ), no effect of Block ( $F[2, 32] = 0.14$ ,  $p = 0.63$ ), and no interaction between these two factors ( $F[2, 32] = 0.611$ ,  $p = 0.54$ ). The non-dominant hand (blue line) moved  $14.6 \pm 1.37$  cm and the head (red line) moved  $21.7 \pm 1.91$  cm.

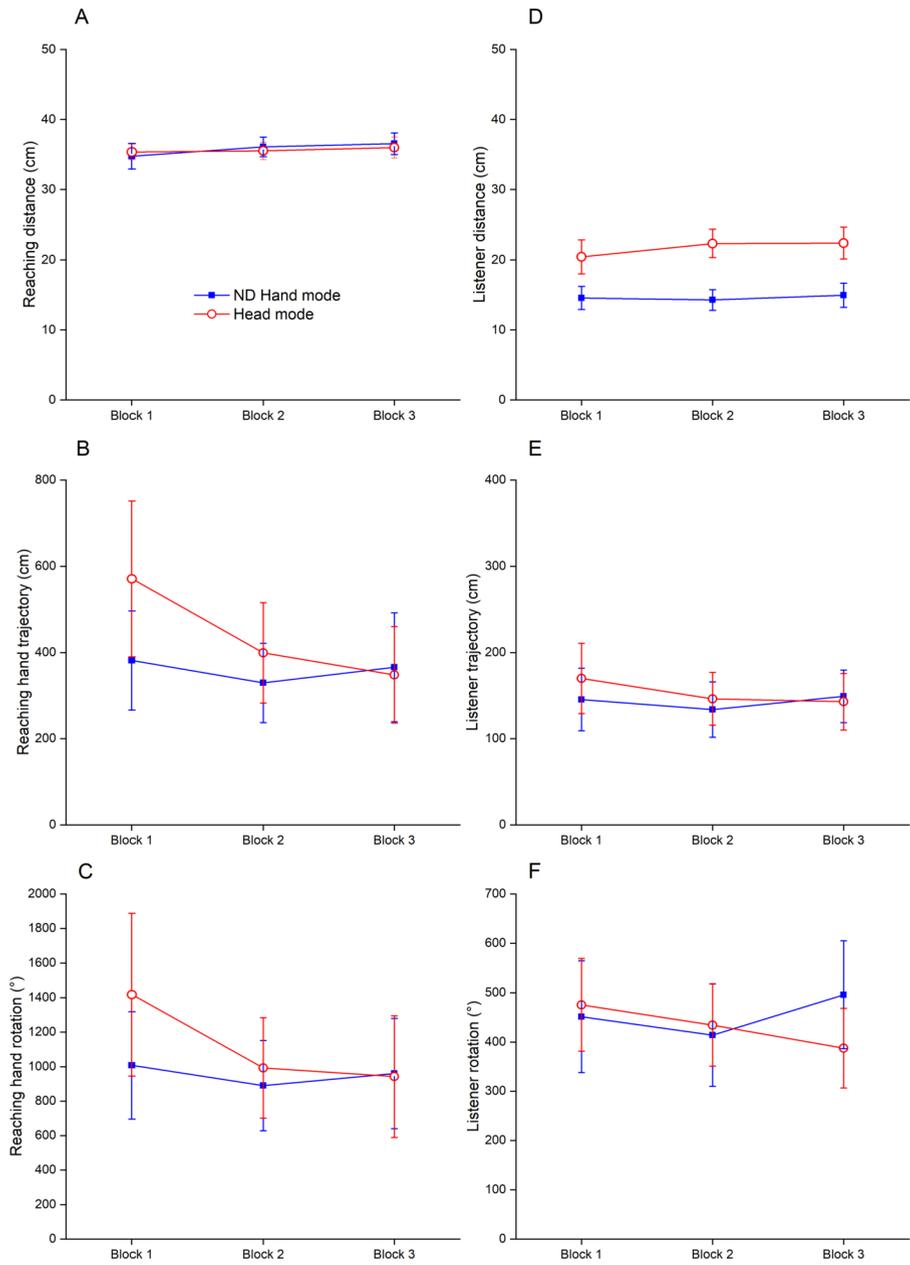
An ANOVA with the factors Mode and Block showed that the listener's cumulative trajectories (Fig. 5E) did not vary with Mode ( $F[1, 16] = 0.11$ ,  $p = 0.74$ ), nor with Block ( $F[2, 32] = 0.24$ ,  $p = 0.79$ ), and there was no interaction ( $F[2, 32] = 0.95$ ,  $p = 0.39$ ). A two factors ANOVA showed that the cumulative rotations of the listener did not vary with Mode ( $F[1, 16] = 1.55$ ,  $p = 0.23$ ) nor with Block ( $F[2, 32] = 0.46$ ,  $p = 0.62$ ), however, there was a significant interaction between Block and Mode ( $F[2, 32] = 3.93$ ,  $p = 0.03$ ). This suggests that the cumulative rotations decrease with Block only when the listener is fixed on the head (Fig. 5F). This interaction was confirmed by Multivariate analysis ( $p = 0.019$ ).

#### 3.2.5. Discussion of Experiment 2

Experiment 2 revealed that the participants were able to adapt their motor behaviour to a novel audio-motor environment that involves a new sensorimotor coordination between the two hands. In this experiment, there was no significant difference between the head mode and the hand mode, neither for the success rate nor for trial duration. These results suggest that the higher performance obtained in the hand mode in Experiment 1 was due to the spatial coincidence between the effector and the sensor. Surprisingly, the head mode that was supposed to be very close to a natural condition (for the human species, the ears are on the head) is not significantly better than the hand mode, and improvements are slower. As in Experiment 1, even if performance seems

Right hand

Listener



to be similar in the two modes, the participants moved their virtual ears with smaller amplitudes in the hand mode.

#### 4. General Discussion

The study reported here investigated adaptation to a novel audio-motor environment. An electromagnetic system recorded the position and orientation of the participants' head and hand and it was connected to a 3D system that provides auditory feedback on movements. The participants' task was to reach an auditory target. The target's sound was perceived by using virtual ears (the listener) located either on the participants' hand (hand mode) or on their head (head mode). In Experiment 1, the participants' hand position was unconstrained and the listener was placed on the same hand as the one used to reach the target. To control for the role of spatial coincidence and movement amplitude, in Experiment 2, hand movements were constrained in order to have similar amplitude and degrees of freedom as those of the head. In addition, the listener was placed on the other hand than the one used to reach the target.

The first main result to emerge from our study was that the participants were able to localize a source within the 3D auditory environment. This result indicates that people are able to learn new efficient motor strategies and to adapt to a new audio-motor environment (see also Safstrom and Edin, 2006). The second main result to emerge from our study was the higher performance obtained when the sensor was located on the hand used to reach the target. This was the case in Experiment 1, in which performance was better in the hand mode than in the head mode. It should be underlined here that in the head mode, the task is mainly a binaural 3D sound source localization task; the reaching component of the task is secondary. The binaural localization of a virtual sound source is critically dependent on the algorithm used for spatialisation. In the present study it was not optimal since it was not personalized by Head Related Transfer Function (HRTF, see Simon *et al.*, 2016). On the other hand, auditory feedback in the hand mode is similar to movement sonification acting as a guide for the reaching hand toward the target (Bevilacqua *et al.*, 2016). In this hand mode, the task resembles more a guided approach task than a pure reaching task. In Experiment 2, the sensor was placed on the other hand and its movements were constrained to be similar to those of the head.

---

**Figure 5.** Kinematics of the reaching hand and of the head in the non-dominant hand mode (in blue) and in the head mode (in red). The left part of the figure shows the kinematic variables of the reaching hand: 3D direct distance (A), cumulative trajectories (B), and cumulative rotations (C). The right part of the figure shows the kinematic variables of the listener: 3D direct distance (D), cumulative trajectories (E), and cumulative rotations (F). The symbols and vertical lines represent the mean and standard error to the mean.

The head mode was supposed to be closer to a natural condition, approximately the position of ears. However, performance was similar in the hand mode and in the head mode and improvements were slower in the head mode. Previous studies have shown that the listener's movements increase both the accuracy of sound localization (Perrett and Noble, 1997; see Kolarik *et al.*, 2016, for a review) and the externalization of sound sources (Brimijoin *et al.*, 2013). This, together with the better control and higher degrees of freedom of hand movements than head movements is a likely explanation of the overall better performance and performance improvement obtained in the hand mode condition.

The better results obtained in the hand mode contradicts some hypotheses from sensorimotor theories of perception according to which the use of a head-mounted camera shares more sensorimotor resemblance with vision than the use of a hand-held camera and thereby might favour better performance (e.g., O'Regan, 2011; O'Regan and Noë, 2001). However, it might be the case that sensorimotor resemblance influences experience on a phenomenological rather than at the behavioural level. It would be interesting to further investigate those factors differentially influencing the behavioural and phenomenological components of the task, as well as the individual differences potentially at play. Indeed, the extent to which the use of a sensory substitution device induces supplemental visual, auditory, tactile, or cognitive processing varies from one user to another (e.g., Auvray *et al.*, 2007a; see also Arnold *et al.*, 2017; Deroy and Auvray, 2014, for discussions on this topic). Thus, in our study, performance appeared to be less linked to the natural aspect related to sensor location than to the spatial coincidence between the body part performing the motor task and the origin of the audio-motor coupling. These results have practical implications, suggesting that under degraded perceptual situations (real or virtual), it could be more efficient to use spatially coincident sensors and effectors.

The third main result from our study concerns the kinematic data, which allow investigating the quality of movement associated with task success. Overall, the kinematics of the reaching hand is not more informative than the success rate. However, the kinematics of the head and non-dominant hand allows a better understanding of the listener's behaviours of active exploration to perform the perceptual task. More precisely, in Experiment 1, the analysis of head movements confirmed that the reaching task was executed differently in the two modes. In the hand mode, the head moved around 10 cm during reaching with relatively small cumulative trajectories and rotations. This displacement can be at least partially explained by the known contribution of the upper trunk to reaching movements (Mark *et al.*, 1997). In contrast, in the head mode, the head roamed the workspace with much larger cumulative trajectories and rotations. The direct 3D displacement of the head at the end

of reaching attained 30 cm which is a strikingly large distance. This observation suggests that in the head mode, the function of the head is not that of a stable egocentric reference frame for planning and executing reaching movement (see also Crawford *et al.*, 2004; Paillard, 1991). Rather, it suggests that head movements, guided by the auditory feedback, are primarily used to explore the 3D workspace in order to ease target detection. We hypothesize that head movements simulate hand movements toward the target. Then, the position and orientation of the head toward the target (heading) can be secondarily used to perform target-directed hand reaching movements (see also Boyer *et al.*, 2013).

Although the two experiments involved differences in terms of constraints on movements and location of the listener, the kinematics of the listener had similarities in the head during head mode in Experiment 1, the head, and the non-dominant hand in Experiment 2, which have similar constraints in terms of amplitude. Indeed, in these three situations, there were large cumulative trajectories and rotations, contrasting with the relatively limited movements of the right reaching hand in hand mode in Experiment 1. In Experiment 2, the kinematics of the head and non-dominant hand are similar. However, the direct distance followed during reaching is smaller when the listener is fixed on the non-dominant hand (around 15 cm) than when it is fixed on the head (around 22 cm). This is quite paradoxical since the movement of the non-dominant hand is free (elbow and wrist degrees of freedom) and probably less effortful than the trunk and neck movements needed to move the head listener. These results suggest that the role of the head to explore space and to indicate target direction (heading) does not transfer easily to the hand.

Participants' performance suggests that the paradigm used in our study can provide useful 3D auditory feedback. This can be applied to a broad range of situations. In the field of sensory substitution, our results are in line with many other studies investigating the relevance of using audition to convey visual information (e.g., the vOICE, Meijer, 1992; Vibe, Hanneton *et al.*, 2010; the Eye-Music, Levy-Tzedek *et al.*, 2014; the See ColOr; Bologna *et al.*, 2009). They additionally suggest that we can step away from reproducing the conditions of natural perception and rather take into account how motor performance relates the best to changes in sensory information. A similar audio-motor coupling could also be used to compensate proprioceptive deficits, for which tactile devices have been mainly used so far (e.g., see Danilov *et al.*, 2007; Diot *et al.*, 2014). In addition, augmented auditory feedback in a game-like situation may be a useful tool for rehabilitation of sensory-motor functions (Schmitz *et al.*, 2018). The hand mode could thus be used as a way of providing augmented feedback to patients with movement disorders such as hemiparesis following stroke. Indeed, due to the strong link between sensory inputs and motor outputs, augmented auditory feedback can be a useful tool in rehabilitation. This

paradigm can also be used to train specific technical skills such as minimally invasive surgery which relies now on virtual reality simulators with enhanced sensory feedback (see Papanikolaou *et al.*, 2019, for a review). Finally, audio-motor feedback can be useful in the field of perceptual augmentation, in which devices are developed to improve access to visual information in a variety of situations, for instance when there is an excess of visual information to process, like when driving a car (e.g., Ho and Spence, 2008) or when conditions of perception are degraded, as can be the case for those involved in firefighting (Carton and Dunne, 2013) or military operations (Jones *et al.*, 2009).

### *Acknowledgements*

We thank Ross Parry for his reading and useful comments on the manuscript. This work was supported by the Mission pour l'Interdisciplinarité (CNRS, Auton, Sublima Grant) and Labex SMART (ANR-11-IDEX-0004-02) supported by French state funds managed by the ANR within the Investissements d'Avenir program under reference ANR-11-IDEX-0004-02.

### **Note**

1. See IA-SIG Interactive 3D Audio Rendering and Evaluation Guidelines (Level 1) and OpenAL Documentation for details: <https://www.openal.org/>.

### **References**

- Ahissar, M. and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning, *Trends Cogn. Sci.* **8**, 457–464.
- Arno, P., Vanlierde, A., Streel, E., Wanet Defalque, M. C., Sanabria-Bohorquez, S. and Veraart, C. (2001). Auditory substitution of vision: pattern recognition by the blind, *Appl. Cogn. Psychol.* **15**, 509–519.
- Arnold, G. and Auvray, M. (2018). Tactile recognition of visual stimuli: specificity *versus* generalization of perceptual learning, *Vision Res.* **152**, 40–50.
- Arnold, G. and Auvray, M. (2014). Perceptual learning: tactile letter recognition transfers across body surfaces, *Multisens. Res.* **27**, 71–90.
- Arnold, G., Pesnot-Lerousseau, J. and Auvray, M. (2017). Individual differences in sensory substitution, *Multisens. Res.* **6**, 579–600.
- Auvray, M., Hanneton, S. and O'Regan, J. K. (2007a). Learning to perceive with a visuo-auditory substitution system: localization and object recognition with 'The vOICe', *Perception* **36**, 416–430.
- Auvray, M. and Myin, E. (2009). Perception with compensatory devices. From sensory substitution to sensorimotor extension, *Cogn. Sci.* **33**, 1036–1058.
- Auvray, M., Philipona, D., O'Regan, J. K. and Spence, C. (2007b). The perception of space and form recognition in a simulated environment: the case of minimalist sensory-substitution devices, *Perception* **36**, 1736–1751.

- Bach-y-Rita, P. and Kercel, S. W. (2003). Sensory substitution and the human-machine interface, *Trends Cogn. Sci.* **7**, 541–546.
- 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.
- Bernstein, N. (1967). *The Coordination and Regulation of Movements*. Pergamon Press, Oxford, UK.
- Bevilacqua, F., Boyer, E. O., Françoise, J., Houix, O., Susini, P., Roby-Brami, A. and Hanneton, S. (2016). Sensori-motor learning with movement sonification: perspectives from recent interdisciplinary studies, *Front. Neurosci.* **10**, 385.
- Bologna, G., Deville, B. and Pun, T. (2009). On the use of the auditory pathway to represent image scenes in real-time, *Neurocomputing* **72**, 839–849.
- Boyer, E. O., Babayan, B. M., Bevilacqua, F., Noisternig, M., Warusfel, O., Roby-Brami, A., Hanneton, S. and Viaud-Delmon, I. (2013). From ear to hand: the role of the auditory-motor loop in pointing to an auditory source, *Front. Comput. Neurosci.* **7**, 26.
- 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.
- Carton, A. and Dunne, L. E. (2013). Tactile distance feedback for firefighters: design and preliminary evaluation of a sensory augmentation glove, in: *Proceedings of the 4th Augmented Human International Conference (AH'13)*, pp. 58–64. ACM, New York, NY, USA.
- Chebat, D. R., Maidenbaum, S. and Amedi, A. (2015). Navigation using sensory substitution in real and virtual mazes, *PLoS ONE* **10**(6), e0126307.
- 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.
- Crawford, J. D., Medendorp, W. P. and Marotta, J. J. (2004). Spatial transformations for eye-hand coordination, *J. Neurophysiol.* **92**, 10–19.
- Danilov, Y., Tyler, M., Skinner, K., Hogle, R. and Bach-y-Rita, P. (2007). Efficacy of electro-tactile vestibular substitution in patients with peripheral and central vestibular loss, *J. Vestib. Res.* **17**, 119–130.
- Deroy, O. and Auvray, M. (2014). A crossmodal perspective on sensory substitution, in: *Perception and Its Modalities*, S. Biggs, M. Matthen and D. Stokes (Eds), pp. 327–349. Oxford University Press, Oxford, UK.
- Diot, B., Halavackova, P., Demongeot, J. and Vuillerme, N. (2014). Sensory substitution for balance control using a vestibular-to-tactile device, *Multisens. Res.* **27**, 313–336.
- Gallay, M., Denis, M. and Auvray, M. (2013). Navigation assistance for blind pedestrians: guidelines for the design of devices and implications for spatial cognition, in: *Representing Space in Cognition: Interrelations of Behaviour, Language, and Formal Models*, T. Tenbrink, J. Wiener and C. Claramunt (Eds), pp. 244–267. Oxford University Press, Oxford, UK.
- 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.
- Hanneton, S., Auvray, M. and Durette, B. (2010). The Vibe: a versatile vision-to-audition sensory substitution device, *Appl. Bionics Biomech.* **7**, 269–276.
- Hanneton, S., Herquel, P. and Auvray, M. (2015). Intermodal recoding of a video game: learning to process signals for motion perception in a pure auditory environment, *Int. J. Adapt. Control Signal Process.* **29**, 1475–1483.

- Ho, C. and Spence, C. (2008). *The Multisensory Driver: Implications for Ergonomic Car Interface Design*. Ashgate Publishing, Aldershot, UK.
- Kim, J. K. and Zatorre, R. J. (2008). Generalized learning of visual-to-auditory substitution in sighted individuals, *Brain Res.* **1242**, 263–275.
- Kim, J. K. and Zatorre, R. J. (2010). Can you hear shapes you touch?, *Exp. Brain Res.* **202**, 747–754.
- Levy-Tzedek, S., Hanassy, S., Abboud, S., Maidenbaum, S. and Amedi, A. (2012). Fast, accurate reaching movements with a visual-to-auditory sensory substitution device, *Restor. Neurol. Neurosci.* **30**, 313–323.
- Levy-Tzedek, S., Riemer, D. and Amedi, A. (2014). Color improves “visual” acuity via sound, *Front. Neurosci.* **8**, 358.
- Mark, L. S., Nemeth, K., Gardner, D., Dainoff, M. J., Paasche, J., Duffy, M. and Grandt, K. (1997). Postural dynamics and the preferred critical boundary for visually guided reaching, *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 1365–1379.
- Marks, D. F. (1973). Visual imagery in the recall of pictures, *Br. J. Psychol.* **64**, 17–24.
- Meijer, P. B. L. (1992). An experimental system for auditory image representations, *IEEE Trans. Bio-Med. Eng.* **39**, 112–121.
- Novak, K. E., Miller, L. E. and Houk, J. C. (2003). Features of motor performance that drive adaptation in rapid hand movements, *Exp. Brain Res.* **148**, 388–400.
- O’Regan, J. K. (2011). *Why Red Doesn’t Sound Like a Bell. Understanding the Feel of Consciousness*. Oxford University Press, Oxford, UK.
- O’Regan, J. K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness, *Behav. Brain Sci.* **24**, 939–973.
- Paillard, J. (1991). Motor and representational framing of space, in: *Brain and Space*, J. Paillard (Ed.), pp. 163–182. Oxford University Press, New York, NY, USA.
- Papanikolaou, I. G., Haidopoulos, D., Paschopoulos, M., Chatzipapas, I., Loutradis, D. and Vlahos, N. F. (2019). Changing the way we train surgeons in the 21st century: a narrative comparative review focused on box trainers and virtual reality simulators, *Eur. J. Obstet. Gynecol. Reprod. Biol.* **235**, 13–18.
- Pollok, B., Schnitzler, I., Mierdorf, T., Stoerig, P. and Schnitzler, A. (2005). Image-to-sound conversion: experience-induced plasticity in auditory cortex of blindfolded adults, *Exp. Brain Res.* **167**, 287–291.
- Proulx, M. J., Brown, D. J., Pasqualotto, A. and Meijer, P. (2014). Multisensory perceptual learning and sensory substitution, *Neurosci. Biobehav. R.* **41**, 16–25.
- 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.
- Ptito, M., Iversen, K., Auvray, M., Deroy, O. and Kupers, R. (2018). Limits of the classical functionalist perspective on sensory substitution, in: *Sensory Substitution and Augmentation*, F. Macpherson (Ed.), *Proceedings of the British Academy*. Oxford University Press, London, UK.
- Rand, M. K., Shimansky, Y., Stelmach, G. E. and Bloedel, J. R. (2004). Adaptation of reach-to-grasp movement in response to force perturbations, *Exp. Brain Res.* **154**, 50–65.
- Renier, L., Laloyaux, C., Collignon, O., Tranduy, D., Vanlierde, A., Bruyer, R. and De Volder, A. G. (2005). The Ponzo illusion using auditory substitution of vision in sighted and early blind subjects, *Perception* **34**, 857–867.

- Safstrom, D. and Edin, B. B. (2006). Acquiring and adapting a novel audiomotor map in human grasping, *Exp. Brain Res.* **173**, 487–497.
- 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.
- Schmitz, G., Bergmann, J., Effenberg, E. O., Krewer, C., Hwang, T. H. and Müller, F. (2018). Movement sonification in stroke rehabilitation, *Front. Neurol.* **9**, 389.
- Simon, L. S., Zacharov, N. and Katz, B. F. (2016). Perceptual attributes for the comparison of head-related transfer functions, *J. Acoust. Soc. Am.* **140**, 3623–3632.
- Striem-Amit, E. and Amedi, A. (2014). Visual cortex extrastriate body-selective area activation in congenitally blind people “seeing” by using sounds, *Curr. Biol.* **6**, 687–692.
- 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.
- Vandenberg, S. G. (1971). *Mental Rotation Test*. University of Colorado, Boulder, CO, USA.