

Flexibility and Stability in Sensory Processing Revealed Using Visual-to-Auditory Sensory Substitution

Uri Hertz^{1,2} and Amir Amedi^{1,2}

¹Department of Medical Neurobiology, Institute for Medical Research Israel-Canada (IMRIC), Hadassah Medical School, Hebrew University of Jerusalem, Jerusalem 91220, Israel and ²Interdisciplinary Center for Neural Computation, The Edmond & Lily Safra Center for Brain Sciences (ELSC), Hebrew University of Jerusalem, Jerusalem 91905, Israel

Address correspondence to Uri Hertz. Email: uri.hertz@mail.huji.ac.il

The classical view of sensory processing involves independent processing in sensory cortices and multisensory integration in associative areas. This hierarchical structure has been challenged by evidence of multisensory responses in sensory areas, and dynamic weighting of sensory inputs in associative areas, thus far reported independently. Here, we used a visual-to-auditory sensory substitution algorithm (SSA) to manipulate the information conveyed by sensory inputs while keeping the stimuli intact. During scan sessions before and after SSA learning, subjects were presented with visual images and auditory soundscapes. The findings reveal 2 dynamic processes. First, crossmodal attenuation of sensory cortices changed direction after SSA learning from visual attenuations of the auditory cortex to auditory attenuations of the visual cortex. Secondly, associative areas changed their sensory response profile from strongest response for visual to that for auditory. The interaction between these phenomena may play an important role in multisensory processing. Consistent features were also found in the sensory dominance in sensory areas and audiovisual convergence in associative area Middle Temporal Gyrus. These 2 factors allow for both stability and a fast, dynamic tuning of the system when required.

Keywords: audiovisual integration, crossmodal effects, fMRI, multisensory processing, sensory cortices

Introduction

Theories of multisensory perception have changed dramatically over the last decade. Traditionally, it was assumed that multiple sensory inputs are processed independently in sensory cortices. Information from multiple senses then converges in associative areas and is then integrated there (Jones and Powell 1970; Zeki 1978; Felleman et al. 1991; Calvert 2001). This rigid, hierarchical structure left little room for crossmodal effects and plasticity. However, behavioral crossmodal effects were reported (Sekuler et al. 1997; Shams et al. 2000), as well as evidence of crossmodal responses in sensory-specific areas (Foxe et al. 2000; Fu et al. 2003; Ghazanfar et al. 2005; Watkins et al. 2006; Kayser et al. 2008). These have weakened the assumption of exclusivity and independence of the sensory cortices (Ghazanfar and Schroeder 2006; Kayser 2010; Shams 2012). However, it is still unclear how the classical hierarchical model and these crossmodal effects interact.

To begin with, evidence of crossmodal responses in sensory cortices is not consistent across studies [there is no doubt, however, that in the blind for instance, V1 shows robust crossmodal responses (Cohen et al. 1997; Bavelier and Neville 2002; Merabet et al. 2005; Pascual-Leone et al. 2005)]. Sensory cortices have been reported to be affected by crossmodal sensory

inputs under a variety of experimental conditions and with different response profiles. Crossmodal enhancements have been found in cases where one sensory input increases the responses in another sensory cortex. For example, the auditory cortex showed an enhanced response when an auditory stimulus was accompanied by visual or tactile stimuli (Foxe et al. 2000, 2002; Fu et al. 2003; Kayser and Logothetis 2007; Lakatos et al. 2007). Neuroimaging studies demonstrated enhancement of auditory cortex responses during audiovisual speech perception (Calvert et al. 1999; Callan, Jones, et al. 2003), and even during silent lip reading (Calvert et al. 1997). In another study, observation of individuals being touched elicited responses in the primary somatosensory cortex (Blake-more et al. 2005). Other studies reported significant and consistent crossmodal attenuations (or deactivations) (Laurienti et al. 2002; Amedi et al. 2005; Mozolic et al. 2008; Kuchinsky et al. 2012). These were usually apparent in situations where sensory inputs compete with one another, for example, when trying to decipher speech in a noisy environment (Kuchinsky et al. 2012), or during visual imagery (Amedi et al. 2005), but were also associated with normal audiovisual speech perception (van Wassenhove et al. 2005). On the other hand, in other studies, crossmodal effects in sensory areas were completely absent, whereas multisensory integrative responses in associative areas were evident (Beauchamp et al. 2004; Noppeney et al. 2010).

In the classic bottom-up, deterministic view of multisensory processing, associative areas are the point of convergence and integration of multiple independent inputs (Jones and Powell 1970; Benevento et al. 1977; Meredith and Stein 1986; Calvert 2001; Beauchamp et al. 2004; van Atteveldt et al. 2004; Stevenson et al. 2009). Recent studies have shown that sensory responses in associative areas can change in a dynamic manner according to the sensory inputs' reliability and saliency (Shams 2012). The Inferior Frontal Sulcus (IFS) and the Intraparietal Sulcus (IPS) were shown to dynamically change their sensory weighting and prefer a more reliable sensory input (Beauchamp et al. 2010; Noppeney et al. 2010). Others have demonstrated how spatial attention may alter sensory responses in the Superior Temporal Sulcus (STS) and Inferior Frontal Gyrus (IFG) (Fairhall and Macaluso 2009). Auditory and visual responses in these areas during audiovisual speech perception were affected both by experience and by the content of the auditory and visual inputs (Lee and Noppeney 2011). This evidence is in line with behavioral results showing dynamic weighting of sensory inputs according to their reliability, resulting in a statistically optimal integration (Ernst and Banks 2002; Fetsch et al. 2009, 2012; Sheppard et al.

2013). In these studies, noise was added to the sensory inputs in order to manipulate their reliability. It is therefore hard to segregate top-down effects such as attention and changes in the information carried by sensory inputs, from bottom-up effects stemming from changes in the sensory inputs. The ways in which these dynamic changes relate to crossmodal effects have yet to be explored.

Here, we examined the hypothesis that the direction and sign of crossmodal effects in sensory cortices and the weighting of sensory inputs in associative areas are both determined by the context in which the sensory inputs are delivered, for example, their information, task relevance, and the novelty of the sensory inputs.

For this purpose, 3 experiments were conducted, in which the experimental context (e.g., task and information conveyed by sensory inputs) was manipulated while stimuli remained the same. This was enabled by using a visual-to-auditory sensory substitution algorithm (SSA), the vOICe (Meijer 1992), which was developed as a rehabilitation tool for the blind. In this SSA, visual information is captured and transformed into auditory soundscapes according to a set of principles. Auditory soundscapes are undecipherable before learning the transformation principles, but become informative after a short learning phase (Kim and Zatorre 2008; Striem-Amit et al. 2012b). Subjects were scanned before and after learning the SSA while perceiving visual images and auditory soundscapes, and while performing an active audiovisual integration task. The stimuli were presented in a semioverlapped manner in which auditory and visual inputs were sometimes given together and sometimes separately, going in and out of synchronization (Hertz and Amedi 2010). This paradigm allows for the detection of auditory and visual responses even when auditory and visual stimuli overlap. Using this paradigm and the SSA manipulation, we were able to detect changes in crossmodal effects in sensory cortices and in sensory responses in associative areas as the information, novelty, and task relevance of the sensory inputs changed.

Methods

Subjects

A total of 12 healthy subjects (5 males and 7 females) aged 22–30 with no neurological deficits were scanned in the current study. The Tel-Aviv Sourasky Medical Center Ethics Committee approved the experimental procedures and written informed consent was obtained from each subject. We rejected the data from one subject because of a technical failure of the auditory system during the scan.

In addition, 14 healthy subjects (7 males and 7 females) aged 24–35 participated in a retinotopy experiment. The Tel-Aviv Sourasky Medical Center Ethics Committee approved the experimental procedure and written informed consent was obtained from each subject.

Stimuli

In this study, visual images were used along with their SSA translation to auditory soundscapes (see Fig. 1C). An SSA used in this study was the “vOICe,” developed by Meijer (1992). The functional basis of this visual-to-auditory transformation lies in a spectrographic sound synthesis from any input image. Time and stereo panning constitute the horizontal axis of the sound representation of an image, tone frequency makes up the vertical axis, and volume corresponds to pixel brightness. Each auditory soundscape displays an image for 1 s as it sweeps from the left side of the image to the right side and usually requires a few repetitions to reconstruct a frame and identify objects. This imposes a serial acquisition of the visual space that differs from the parallel nature of visual acquisition of information from images in

which the entire image is available at once. To make the visual images similar to the auditory soundscapes, we did not present the entire image at once, but rather had a mask sweep across the image from left to right for 1 s, revealing the image a little at a time, similar to a vector spotlight sweeping across an image in the soundscapes.

Learning SSA

All subjects experienced the “vOICe” sensory substitution for the first time only in the middle of the functional magnetic resonance imaging (fMRI) session (see Fig. 1A) and were completely naive to the principles of the “vOICe,” described above, before the learning session. During the brief learning phase, subjects were first instructed about the visual-to-auditory transformation principles, and then proceeded to practice very simple shape and location perception of a standardized set of stimuli that is part of the training set of stimuli used in our laboratory to teach blind individuals to use the “vOICe” (including small lines, rectangles, and round objects presented at 4 possible locations on the screen). Feedback regarding performance was given by the instructor. This procedure has proved to be efficient in triggering recruitment of both the ventral and dorsal pathways by the soundscapes in both the sighted and the blind while performing simple shape versus localization tasks, respectively (Striem-Amit et al. 2012b; see the same reference for a further description of this 1–1.5 h training session).

Experimental Design

This study consisted of 3 experimental conditions with a short (1 h) learning session in the middle (Fig. 1A). All experiments included blocks of auditory stimuli and blocks of visual stimuli, delivered with different presentation rates (e.g., auditory blocks repetitions and visual blocks repetitions were not the same), in a semioverlapped manner (Calvert et al. 2000; Hertz and Amedi 2010). This means that sometimes auditory and visual stimuli were presented at the same time, sometimes only auditory or only visual stimuli were presented, and sometimes no stimuli were presented at all (Fig. 1B). This design allowed for detection of auditory and visual responses even when they were presented together [see General Linear Model (GLM) description below; Hertz and Amedi 2010]. In the first experiment (“Pre”), blocks of visual images repeated 21 times, whereas auditory blocks repeated 14 times (Fig. 1B). Visual blocks included 3 shapes (a circle, a horizontal line, and a staircase); each was repeated 4 times, totaling 12 s per block. Auditory blocks included 3 soundscapes, which were the SSA translation of a circle, a horizontal line, and a staircase. Each soundscape repeated 4 times, totaling 12 s per block. The rest was 15 s between auditory blocks, and 6 s between visual blocks. For this condition, we chose a passive paradigm in which subjects were instructed to maintain fixation on a red cross in the middle of the screen and passively attend to the stimuli. This was done to minimize task-related effects, so that the changes in cortical activity could be assigned to the changes in the information conveyed by auditory input following learning (Calvert et al. 2000; van Atteveldt et al. 2004; Naumer et al. 2009).

For the second condition, which was conducted after learning (“Plus”), we used an active paradigm which forced subjects to attend to both the auditory and visual inputs. In this experiment, auditory soundscape blocks and visual image blocks were presented to the subjects, auditory blocks repeated 20 times, and visual blocks repeated 15 times (Fig. 1D). Auditory blocks included 4 soundscapes, each repeated 3 times, totaling 12 s per auditory block, which was followed by 6 s of rest. Visual blocks included 6 images, each repeated 3 times, totaling 18 s per visual block, which was followed by 6 s of rest. Subjects were instructed to press a button when they perceived a combination of a vertical line and a horizontal line, one via a visual image and the other via an auditory soundscape that when combined formed a multisensory “Plus” (+) sign (Fig. 1D in green dashed lines). These “Plus” events occurred 10 times during this experiment as shown in the green rectangle in Figure 1D.

The final experiment condition was a repetition of the “Pre-Passive” experiment described above, after learning SSA (“Post”). The same visual and auditory blocks as in the “Pre” experiment were delivered, and no active response was required. Subjects were instructed to

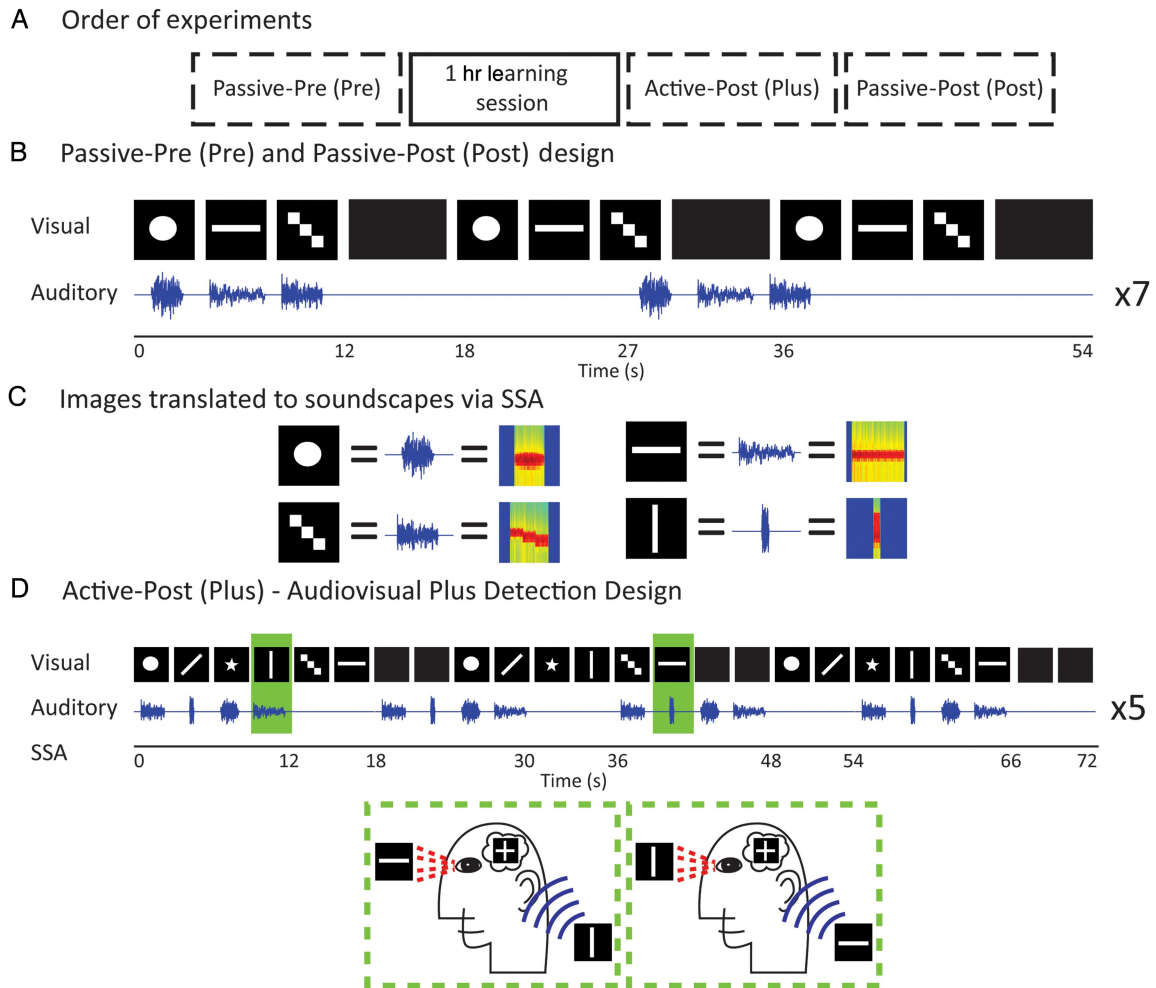


Figure 1. Order and design of experimental sessions. (A) Order of experiments—this study consisted of 3 sessions of data acquisition, interspersed by a short (1 h) learning session outside the scanner. (B) Passive paradigm—visual images and auditory soundscapes were presented to the subjects in a semioverlapped manner, while subjects were asked to maintain fixation on a cross in the middle of the screen. Each visual block contained 3 images, each presented for 4 s, totaling 12 s per block. Each auditory block contained 3 soundscapes, each was 1 s long and repeated 4 times, totaling 12 s for block. Auditory and visual blocks had different representation rates, as 3 visual blocks were presented for every 2 auditory blocks, totally auditory blocks were repeated 14 times throughout the experiment, and visual blocks were repeated 21 times. This experiment was carried out once before the learning period (Passive–Pre), and once after learning (Passive–Post). (C) Visual images and their translation to soundscapes via SSA. The soundscapes used throughout the experiments represented 4 visual images. Their translation to sound is represented by the time course of the sound (blue line) and by spectrograms of the soundscape (*y*-axis depicts frequency, *x*-axis depicts time, red colors are high energy and blues are low energy). This translation is not trivial and was not deciphered before learning SSA. Shape information can only be extracted from the soundscapes after learning (spectrograms reveal the shape information encoded in the soundscapes). (D) Active Post–Audiovisual plus detection. The subjects were instructed to press a button when they perceived a combination of a vertical line and a horizontal line, each from a different modality, either auditory or visual, that formed a multisensory plus (+) sign (demonstrated in the green dashed boxes). Auditory soundscape blocks and visual image blocks were presented to the subjects in a semioverlapped manner. Auditory blocks lasted 12 s, containing four 1 s soundscapes, each repeated 3 times. Visual blocks lasted 18 s and included 6 images. The auditory blocks were repeated 20 times throughout the experiment, and the visual blocks repeated 15 times. “Plus” events occurred 10 times throughout the experiment (these events are marked in green).

maintain fixation on a red cross in the middle of the screen and passively attend to the stimuli.

Functional and Anatomical MRI Acquisition

The blood oxygen level-dependent (BOLD) fMRI measurements were conducted with a GE 3-T echo planar imaging system. All images were acquired using a standard quadrature head coil. The scanning session included anatomical and functional imaging. Three-dimensional anatomical volumes were collected using a T_1 spoiled gradient echo sequence. Functional data were obtained under the following timing parameters: TR (time repetition) = 1.5 s, TE (time echo) = 30 ms, flip angle (FA) = 70°, imaging matrix = 64 × 64, field of view (FOV) = 20 × 20 cm. Twenty-nine slices with slice thickness = 4 mm and no gap were oriented in the axial position for complete coverage of the whole cortex and scanned in an interleaved order. The first 5 images (during the first baseline rest condition) were excluded from the analysis because of nonsteady-state magnetization.

Retinotopy data were acquired using a whole-body, 3-T Magnetom Trio scanner (Siemens, Germany). The fMRI protocols were based on multislice gradient echoplanar imaging and a standard head coil. The functional data were collected under the following timing parameters: TR = 1.5 s, TE = 30 ms, FA = 70°, imaging matrix = 80 × 80, FOV = 24 × 24 cm (i.e., in-plane resolution of 3 mm). Twenty-two slices with slice thickness = 4.5 mm and 0.5 mm gap were oriented in the axial position for complete coverage of the whole cortex.

Cortical reconstruction included the segmentation of the white matter by using a grow-region function embedded in the Brain Voyager QX 2.0.8 software package. The cortical surface was then inflated.

Preprocessing of fMRI Data

Data analysis was initially performed using the Brain Voyager QX 2.0.8 software package (Brain Innovation, Maastricht, Netherlands). For spectral analysis, fMRI data went through several preprocessing steps which included head motion correction, slice scan time correction, and

high-pass filtering (cutoff frequency: 3 cycles/scan) using temporal smoothing (4 s) in the frequency domain to remove drifts and to improve the signal-to-noise ratio. No data included in the study showed translational motion exceeding 2 mm in any given axis, or had spike-like motion of >1 mm in any direction. Functional data also underwent spatial smoothing (spatial Gaussian smoothing, full width at half maximum = 6 mm). The time courses were de-trended to remove linear drifts. Additionally, the time courses were normalized using z-scores to allow comparison of beta values between subjects. Functional and anatomical datasets for each subject were aligned and fit to the standardized Talairach space.

GLM and Analysis of Variance

One GLM was constructed to analyze the learning effect (“Pre” and “Post”) and the task effect (“Plus”). The model included predictors from all 3 experimental conditions for all the subjects, in a repeated-measures design. Predictors were created by using a step function for the selected condition and convolving it with a hemodynamic response function (HRF; (Boynton et al. 1996). Auditory and visual predictors for each experimental condition were created in this manner (e.g., “Auditory-Pre,” “Auditory-Post,” and “Auditory-Plus,” in the auditory case). Our semioverlapped design means that auditory and visual predictors overlapped in some intervals, but remained orthogonal because the auditory and visual stimulus presentation rates were different (in the passive experiments 14 repetitions and 21 repetitions, respectively), and thus did not violate the nonsingularity of the design matrix (Friston, Holmes, et al. 1994). In the passive experiments, other predictors were used in this GLM to model the audiovisual interactions, in which only synchronized audiovisual trials were convolved with the HRF. Auditory, visual, and interaction predictors could be alternatively described as an approximation of the 3 Fourier coefficients of the time courses in the audiovisual interaction representation frequency, and the auditory and visual stimulus representation frequencies (Hertz and Amedi 2010). This design was previously used to detect tonotopic and retinotopic maps in the same experiment and successfully detected the auditory and visual contributions to voxel responses even when they were presented in the same time (Hertz and Amedi 2010). The model also contained predictors for the plus events (light green events in Fig. 1D) in the “Plus” experiment, which also had a different representation rate from the auditory and visual stimuli (10 repetitions). As in standard GLM analysis, the model also included a constant predictor to estimate the constant baseline activity (Friston, Jezzard, et al. 1994).

The GLM was applied to the entire group of subjects in a hierarchical random-effects analysis (RFX; Friston et al. 1999); see, for instance, implementation in Amedi et al. (2007). Beta values estimated for the visual and auditory predictors in the GLM of the passive pre–post sessions were analyzed using two-way analysis of variance (ANOVA) (Modality [Auditory, Visual] × Learning [Pre, Post]). Main effects and interaction effects were examined, and post hoc contrasts were used to determine the direction of the effect. The resulting contrast maps were set to a threshold of $P < 0.05$, and then corrected for multiple comparisons using the cluster size correction plugin of BrainVoyager (Forman et al. 1995). Auditory and visual beta values were also used to examine group auditory and visual responses, and were compared with zero to find significant responses (Friston et al. 1999). It should be noted that negative beta values do not necessarily mean negative BOLD signal because auditory and visual predictors were semioverlapped. For example, visual areas may have a positive response for visual stimuli, but when accompanied by auditory stimuli (both are delivered in the same time) the positive response might be attenuated, though still positive. In this case, the auditory predictor is assigned a negative response, as it attenuates the activation in the voxel, without the BOLD signal dropping below the average response.

Tonotopic and Retinotopic Maps

Our experiments were aimed at determining the dynamic nature of sensory responses; for example, how they change according to the experimental context in which they are delivered. We decided to use

independent datasets, not confounded by our experimental design or the group of subjects who participated in the main experiment, and to provide an approximation of the boundaries of sensory cortices to which our main results could be compared. We delineated sensory cortices according to their response to pure tones and retinotopic stimuli (Serenio et al. 1995; Engel et al. 1997; Striem-Amit et al. 2011). Data from 2 independent experiments were used here to establish these boundaries. None of the 3 groups of subjects (retinotopy, tonotopy, and main experiment) overlapped. The boundaries of the auditory cortex were estimated using the tonotopy data from Striem-Amit et al. 2011. In their experiments, 10 subjects were scanned while listening to an ascending 30-s chirp, ranging from 250 Hz to 4 kHz, repeated 15 times (a descending chirp was also used in the original experiment to verify the results of the rising chirp). First, tonotopic responsive areas were detected as areas with a significant response to auditory stimuli. The tonotopic organization within these areas, for example, multiple gradual maps of preferred tones (from high to low and vice versa), was detected. These were found in core auditory areas and in associative auditory area, extending toward the temporal areas. Average maps from these 10 subjects were used here to determine cortical areas responsive to pure tones ($P < 0.05$, corr.) and to delineate boundaries between multiple tonotopic maps (delineated on a flattened cortical reconstruction in Fig. 2). It should be noted that the tonotopic responsive areas include associative auditory areas (such as the Superior Temporal Gyrus) and are not confounds of the auditory core areas.

Retinotopic organization was detected using a polar angle experiment (Serenio et al. 1995; Engel et al. 1997). The visual stimuli were adapted from standard retinotopy mapping using a rotating wedge with a polar angle of 22.5°. The wedge rotated around a fixation point 20 times, completing a full cycle every 30 s. It contained a flickering (6 Hz) radial checkerboard pattern according to standard retinotopic procedures. The stimuli were projected via an LCD projector onto a tangent screen positioned over the subject's forehead and viewed through a tilted mirror. Here, we detected retinotopic responsive areas, which were defined as areas with a significant response to visual stimuli. The retinotopic organization within these areas followed, as multiple gradual maps of preferred polar angles were detected. Average maps from 14 subjects were used to determine cortical areas responsive to retinotopic stimuli ($P < 0.05$, corr.) and to delineate boundaries between multiple retinotopic maps (delineated on a flattened cortical reconstruction in Fig. 2; see Hertz and Amedi (2010) for a detailed description). It should be noted that the retinotopic responsive areas include high sensory-specific areas (like V3 and V4) and are not confounds of the Calcarine Sulcus and V1.

Results

Three experiments were carried out, all of which involved visual images and auditory soundscapes (translations of visual images to sound through an SSA) that were delivered in the same experiment in a semioverlapping manner (Fig. 1, see Experimental Procedures section for detailed description). During the first experimental condition, subjects were scanned while passively perceiving the visual and auditory stimuli before learning the visual-to-auditory SSA transformation (“Pre,” Fig. 1B). This scan was followed by a brief, 1-h learning session, in which the principles of the SSA were introduced and demonstrated. Afterwards, the subjects were scanned again in 2 experimental conditions. In the first condition, the subjects were asked to press a button when they detected an audiovisual plus sign, which was constructed from a combination of vertical and horizontal lines delivered as an auditory soundscape or a visual image (“Plus,” Fig. 1D). This task was designed to be very easy to perform (average success rate was $96 \pm 6.5\%$). Nine of 12 subjects achieved 100% success rate, 90% for 2 subjects, and 80% for 1 subject, but enforced integration of visual information and SSA auditory information to

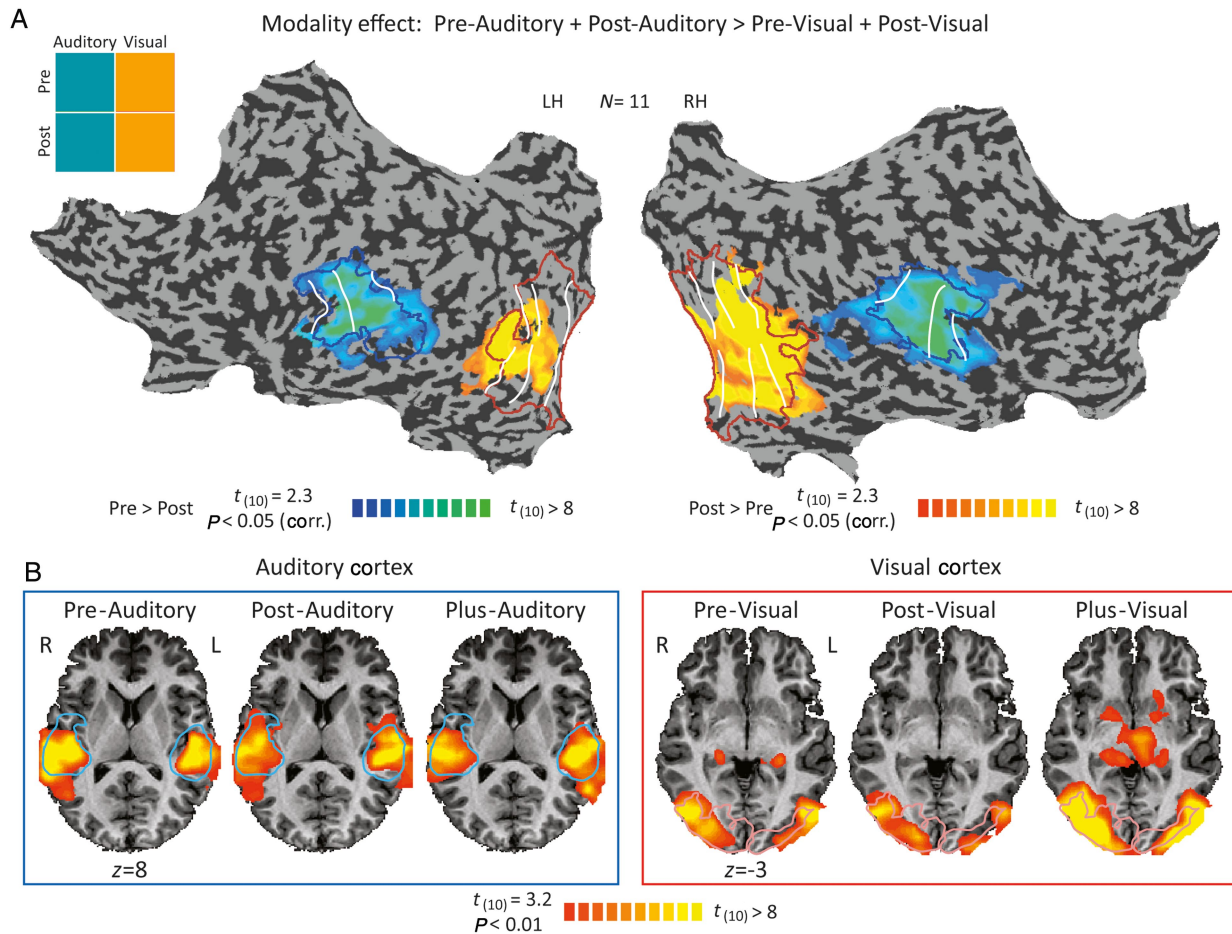


Figure 2. Consistent sensory preference in sensory areas. (A) Statistical parametric map of the modality effect revealed by a two-way ANOVA, vision versus auditory preference ($P < 0.05$, corr.), is presented on a flattened cortical reconstruction of one of the subjects. The analysis was carried out within areas that were responsive to either vision or auditory, before or after learning SSA. Auditory and visual responses are distinct and localized in accordance with their respective sensory areas, as defined by retinotopic (red) and tonotopic blue borders. This was enabled even though the auditory and visual stimuli were delivered at the same time, in a semioverlapped manner (see Methods). White lines delineate group average of borders between tonotopic and retinotopic gradients. (B) Auditory responses in the auditory cortex (blue box, left) and visual responses in the visual cortex (red box, right) in all 3 experimental conditions (Pre, Post, and Plus, in left to right panels), compared with baseline. Auditory responses in auditory cortex were always positive ($P < 0.01$, uncorrected), as were the visual responses in the visual cortex ($P < 0.01$, uncorrected), also demonstrated by beta values in Figure 2B. Here, as well tonotopic borders are in blue and retinotopic borders are depicted in red.

correctly identify the specific audiovisual combinations. Then, another passive experiment was carried out (“Post,” Fig. 1B), which was exactly the same as the pre-learning session in terms of stimuli and task, but the auditory stimuli took on a different meaning as a result of the SSA learning session.

The fact that visual and auditory stimuli were presented in the same experiment in a semioverlapped manner in which auditory soundscapes and visual images had different presentation rates enabled detection of auditory and visual responses separately using a method we developed recently (Hertz and Amedi 2010). Auditory, visual, and interaction predictors from all 3 experiments (“Pre,” “Post,” and “Plus”) were evaluated in one GLM, resulting in beta values for all these conditions for all the subjects (see Methods). Auditory and visual beta values from the 2 passive experiments were analyzed using a two-way ANOVA (Modality [Auditory, Visual] \times Learning [Pre, Post]). Main effects (e.g., Modality and Learning effect) and interaction effects were examined, and post hoc contrasts were used to determine the direction of the effect (see Table 1). Auditory and visual responses in the “Plus” experimental condition were tested within regions of interest, defined by the

ANOVA results from the passive experiments, as well as the overall visual and auditory responses in this condition (see Table 2).

Modality Effect

To examine the consistent sensory preference of sensory areas, the ANOVA modality effect was used. A statistical parametric map of the ANOVA modality effect was thresholded at $F_{1,10} = 6$ ($P < 0.05$) (Table 1, modality effect). The largest clusters were on the left and right Planum Temporale (PT), and the left and right occipital pole. A post hoc contrast between auditory and visual conditions was carried out to determine the direction of the effect (Fig. 2A). Areas which did not show positive responses to either auditory or visual stimuli were removed from the map, and it was corrected for multiple comparisons (Forman et al. 1995). This map revealed consistent sensory responses in sensory areas; such as, auditory positive responses in auditory areas and visual-positive responses in visual areas. This map was congruent with the delineation of tonotopic and retinotopic responsive areas (see Methods for

Table 1
ANOVA main effects and interaction effects

Name	Laterality	Brodman	x	y	z	Number of voxels
Modality effect ($F_{1,10} > 6, P < 0.05$, uncorrected)						
Middle Occipital Gyrus	L	9	31	-78	-7	52 605
Planum Temporale	L	40	51	-24	10	46 290
Fusiform Gyrus	L	46	-39	-81	-10	34 769
Planum Temporale	R	22	-53	-27	10	29 006
Precuneus	R	2	18	-76	46	4892
Thalamus	L + R	6	17	-33	1	1177
Precentral Gyrus	R	18	53	-8	44	1060
Angular Gyrus	R	6	-46	-60	33	1026
Thalamus	R	13	-23	-33	-2	870
Superior Frontal Gyrus	R	18	-33	63	4	774
Inferior Parietal Sulcus	L	22	-51	-32	36	386
Learning effect ($F_{1,10} > 6, P < 0.05$, uncorrected)						
Inferior Occipital Gyrus	L	19	-44	-79	-4	2507
Fusiform Gyrus	R	37	44	-58	-9	888
Middle Frontal Gyrus	R	9	43	9	29	825
Planum Temporale	R	41	43	-10	8	776
Planum Temporale	L	41	-44	-12	7	396
Precuneus	R	7	24	-67	39	351
Interaction effect ($F_{1,10} > 6, P < 0.05$, uncorrected)						
Inferior Frontal Gyrus	L	9	-47	4	23	15 990
Supramarginal Gyrus	L	40	-56	-37	35	13 989
Middle Frontal Gyrus	L	46	-48	37	17	5937
Superior Temporal Gyrus	R	22	51	-17	0	3264
Postcentral Gyrus	R	2	59	-24	37	3165
Superior Frontal Gyrus	L + R	6	-2	2	53	2822
Cuneus	R	18	24	-87	11	2702
Prefrontal Gyrus	R	6	39	-5	51	1737
Insula	R	13	40	3	19	1701
lingual Gyrus	R	18	13	-87	-11	844
Middle Temporal Gyrus	L	22	-60	-30	1	460
Superior Temporal Gyrus	L	22	-58	-12	-1	420
Lentiform	R		14	-1	6	352
Precuneus	R	7	16	-69	48	332

Note: Maps were thresholded at $F_{1,10} > 6$ ($P < 0.05$), not corrected for multiple comparisons. Clusters that contained fewer than 300 voxels are not displayed.

details about the tonotopic and retinotopic maps and acquisition). This faithful detection of auditory and visual responses despite the fact that they were administered together strengthened the reliability of the analysis and design used in this study. These maps were not exactly concordant, however, because our visual stimuli did not elicit a great deal of activity in the periphery responsive areas, and our auditory stimuli elicited responses in associative auditory areas beyond the boundaries determined by the tonotopy experiment. Sensory areas were therefore sensory specific to a large extent, in fitting with the textbook canonic traditional view of the brain (Adrian 1949; Felleman et al. 1991). This agreement verifies our experimental design and analysis. These findings will serve as a baseline for further analyses.

Auditory and visual responses in all 3 experiments were examined within sensory-specific areas. The statistical parametric maps describe significant responses to either the auditory or the visual condition (significant difference from zero, see Methods; Fig. 2B). Positive responses were found in the visual cortex in all visual conditions (“Visual-Pre,” “Visual-Post,” and “Visual-Plus”; see also Tables 1 and 2). Positive responses in the auditory cortex were found in all auditory conditions (“Auditory-Pre,” “Auditory-Post,” and “Auditory-Plus”; see also Tables 1 and 2).

Learning Effect

We used the ANOVA learning effect to chart the changes in sensory responses following learning (Table 1, learning

effect). Post hoc analysis was carried out to examine the direction of the effect, and post-learning conditions (auditory and visual) were compared with the pre-learning conditions (Fig. 3A). The learning effect was examined within areas which were positively responsive to either auditory or visual stimuli. A significant preference for post-learning conditions was found in the right PT (Brodmann 41), and a significant preference for the pre-learning conditions was found in the left Inferior Occipital Gyrus (IOG, Brodmann 19; Fig. 3A, $P < 0.05$, corr.). The positive cluster in the auditory cortex was located well within the areas responsive to auditory stimuli as defined by the ANOVA modality effect (Fig. 2A, in blue), and was on the border of the tonotopic responsive areas with some of it within and some outside. This indicates that the learning effect was not only in the primary auditory area, but also in the associative auditory areas (Striem-Amit et al. 2011), in line with previous crossmodal effect reports in the auditory cortex (Foxe et al. 2002; Kayser et al. 2008). The visual cluster lay within the retinotopic boundaries in the foveal part of area V3 and IO. To better understand these learning effects, the GLM-evaluated beta values were sampled from these areas (Fig. 3B). Auditory and visual beta values from the 3 experimental conditions (“Pre,” “Post,” and “Plus”) were tested for significant responses, and the results were corrected for multiple comparisons using a false discovery rate (FDR) correction. In the left IOG, positive beta values were found for visual conditions in all 3 experiments. However, the auditory responses were not significantly different from zero before learning and were significantly negative in the “Post” condition. This suggests that the preference for the pre-learning condition found in visual areas was due to the auditory attenuations (as suggested by the negative beta values) in this area after learning the visual-to-auditory SSA. It should be noted that negative beta values do not necessarily mean a negative BOLD signal, but rather attenuations of the positive signal. In this case, visual responses in visual areas were significantly lower, yet still positive, when accompanied by auditory stimuli after SSA learning, compared with times when those were not accompanied by auditory stimuli (see Methods for further details).

The reverse pattern was found in the auditory cortex. In the right PT, auditory responses were significantly positive in all 3 experimental conditions. However, visual responses were negative in the “Passive-Pre” condition, before learning SSA, and not significantly different from zero in the “Passive-Post” condition. This suggests that, similar to the logic detailed above, the preference for post-learning conditions in this area was due to its release from visual attenuations (as suggested by the negative beta values) after learning the SSA.

These crossmodal effects were also found during auditory and visual responses in all 3 experiments. The statistical parametric maps describe significant responses to either the auditory or the visual condition (significant difference from zero, see Methods) (Fig. 3C). Negative auditory responses were found in visual areas after learning in the “Post” experiment, but not before learning. In the auditory cortex, negative visual responses were found before learning, but not in the “Post” experiment.

Crossmodal effects were also examined in the “Plus” experiment. The beta values for auditory and visual responses during the “Plus” experiment were sampled from the same area as the “Pre” and “Post” beta values, defined by the significant clusters of the learning effect (Fig. 3B). Crossmodal attenuations were found in the auditory cortex, in that the visual responses during

Table 2

Auditory and visual responses in the “Plus” condition

Name	Laterality	Brodman	x	y	z	Number of voxels
Plus—auditory						
Positive ($t_{(10)} > 2.5, P < 0.05$, uncorrected)						
Planum Temporale	L	42	-53	-33	10	38 409
Intra Parietal Sulcus						
Superior Temporal Sulcus						
Planum Temporale	R	41	52	-29	8	25 589
Superior Temporal Sulcus						
Middle Temporal Gyrus						
Inferior Frontal Sulcus	L	44	-47	10	28	10 745
Precentral Gyrus	L	6	-38	-7	56	811
Supplementary Motor Area	L + R	6	-3	7	51	699
Negative ($t_{(10)} < -2.5, P < 0.05$, uncorrected)						
Medial Frontal Cortex	L + R	32	5	46	9	23 605
Middle Occipital Gyrus	L	19	-23	-85	5	6729
Cingulate Gyrus	L	31	-21	-31	44	5747
Central Sulcus	R	6	14	-25	54	3438
Calcarine Sulcus	R	18	10	-77	-6	1202
Parahippocampal Gyrus	R	37	33	-33	-9	1186
Plus—visual						
Positive ($t_{(10)} > 2.5, P < 0.05$, uncorrected)						
Occipital Cortex	L + R	17, 18, 19, 7, 23	2	-63	14	182 745
Intraparietal Sulcus						
Thalamus	L + R		-2	-15	2	25 172
Superior Colliculus						
Inferior Frontal Gyrus	L	9	-46	3	32	7868
Precentral Gyrus	R	6	29	-9	52	2780
Supplementary Motor Area	L + R	6	-3	1	51	2103
Precentral Gyrus	L	6	-29	-12	52	1430
Middle Frontal Gyrus	R	9	46	25	26	1162
Negative ($t_{(10)} < -2.5, P < 0.05$, uncorrected)						
Anterior Cingulate Gyrus	L + R	32	4	35	11	1340
Superior Temporal Gyrus	R	22	48	-18	-2	900
Cingulate Gyrus	L	31	-12	-41	37	871
Angular Gyrus	R	39	49	-66	27	527
Superior Temporal Gyrus	L	22	-57	-33	2	336

the “Plus” experiment were negative. A trend toward a negative auditory response in the visual cortex was also found. To further explore the crossmodal effects during the “Plus” experiment, the beta values of “Plus-Auditory” and “Plus-Visual” conditions were statistically assessed (Fig. 3C and Table 2). These revealed crossmodal attenuations in both the auditory and visual cortex ($P < 0.01$), for example, negative visual responses were detected in the auditory cortex, and negative auditory responses were detected in the visual cortex.

The changes in the pattern of negative crossmodal responses seem to follow the changes in information conveyed by the sensory inputs, their novelty, and task relevance. Before SSA learning, the auditory inputs are undecipherable, and visual images are the only informative inputs. It is in this condition that the auditory cortex undergoes attenuation whenever a visual stimulus is shown. After SSA learning, auditory soundscapes become informative and novel, and attenuate visual cortex responses. When both inputs are task relevant, crossmodal attenuation of both sensory areas is seen.

Thus, overall, specific sensory areas are characterized by consistent and stable unisensory dominance, regardless of context and task, which is congruent with the classic bottom-up view of sensory perception. These consistent sensory responses are nevertheless attenuated by crossmodal effects, in a context-dependent manner; namely, noninformative input is inhibited by the informative stimuli (before SSA learning), and once it becomes informative the direction of the crossmodal effect is reversed from the newly learned sensory input to the

veteran sensory input. Interestingly, in the “Plus” experiment, when both sensory inputs were task relevant, crossmodal attenuations were found in both sensory cortices.

Interaction Effect—Shift in Sensory Responses’ Profiles

Importantly, the ANOVA interaction effect (Modality \times Learning) revealed a dynamic shift in sensory preference in associative areas (Table 1, interaction effect). These areas responded to both sensory inputs, but changed their sensory response profile, that is, which sensory input elicited the strongest response (which sensory input was preferred). A post hoc contrast was carried out to detect the direction of the sensory shift (Fig. 4A). This comparison was carried out within areas which were responsive either to auditory or visual inputs, before or after learning (i.e., these areas had to exhibit a positive response to at least one of the sensory conditions, in one of the experiments), and was corrected for cluster size. Only one direction of sensory preference shift was found: from visual preference to auditory preference; namely, visual responses were higher than auditory responses before SSA learning, and auditory responses were stronger than visual responses after SSA learning. This change in the profile of sensory responses does not mean that these areas responded to only one sensory input, but that the relation between the strength of the auditory and visual responses changed. This direction of sensory preference shift is in line with the fact that learning involved the auditory input, by transforming it from noninformative to

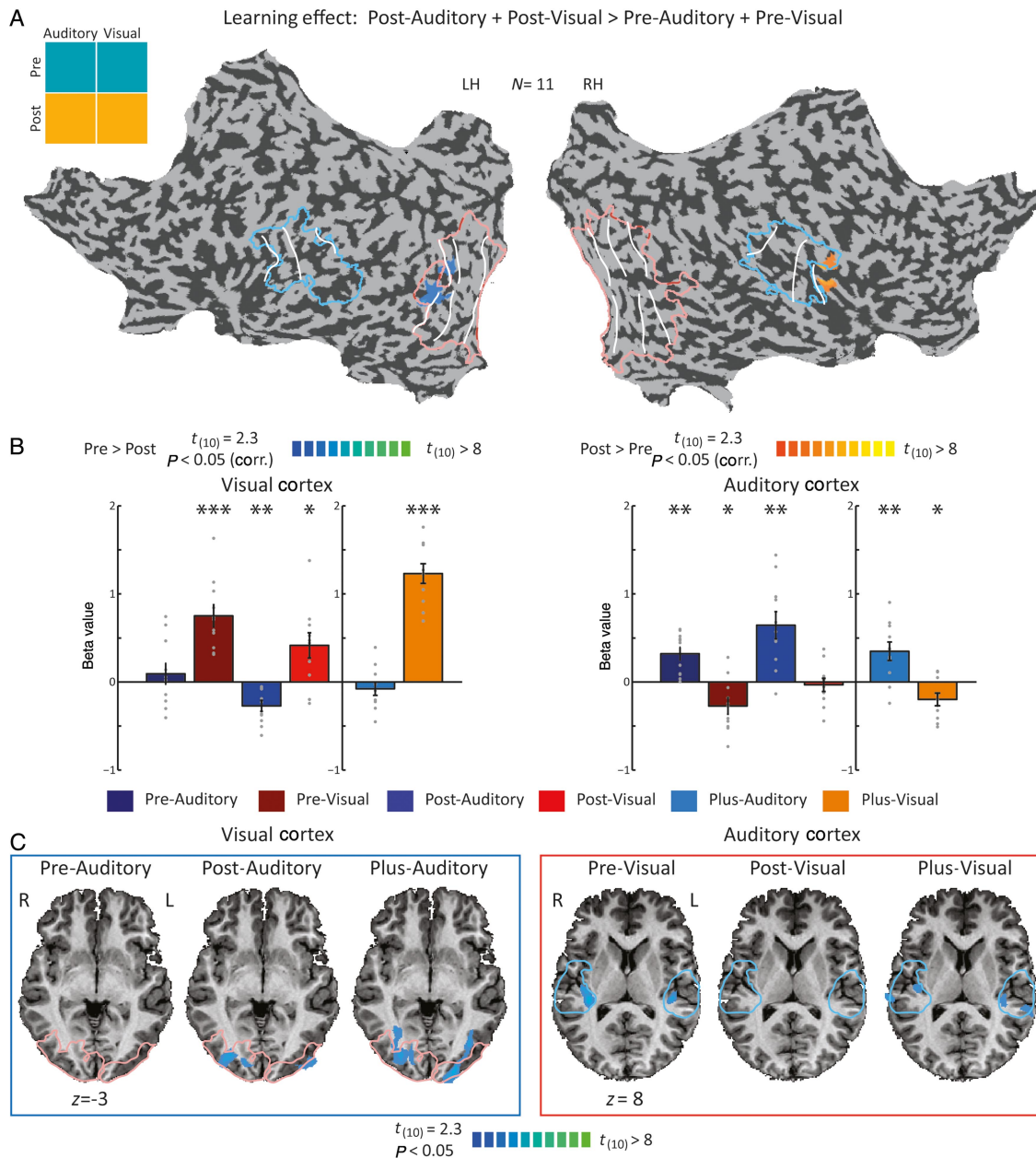


Figure 3. Dynamic crossmodal attenuations of sensory areas. (A) Statistical parametric map of the learning effect revealed by a two-way ANOVA, post-learning versus pre-learning preference ($P < 0.05$, corr.), is presented on a flattened cortical reconstruction of one of the subjects. The analysis was carried out within areas that were responsive to either vision or auditory inputs, before or after learning SSA. The statistical parametric map revealed a preference for the pre-learning condition within the right visual cortex, and a preference for the post-learning condition within the left auditory cortex. Retinotopic and tonotopic borders are presented as well; blue lines delineate tonotopic areas and red lines delineate retinotopic areas. (B) Beta values sampled from the clusters depicted in (A) reveal that the learning effect stemmed from changes in crossmodal attenuations (grey dots represent single subjects' beta values; means and SD are presented). In primary visual areas (on the left), visual responses were significantly positive throughout the experiments ($*P < 0.05$, $**P < 0.005$, $***P < 0.0005$, corr.). However, auditory responses were significantly negative during the Post-passive experiment, underlying the learning effect in this area. In the primary auditory cluster, auditory stimuli elicited significant positive responses throughout the experiments, but visual responses were negative in the “Pre” and “Plus” experiments. (C) Auditory responses in the visual cortex (blue box, left) and visual responses in the auditory cortex (red box, right) in all 3 experimental conditions (Pre, Post, and Plus, in left to right panels), compared with baseline. In all cases, only negative responses were detected, if any ($P < 0.05$, uncorrected). Here, as well tonotopic borders are in blue and retinotopic borders are depicted in red. In the visual cortex, auditory responses were not presented before learning SSA, but appeared both after learning and in the audiovisual integration task. This experimental context-dependent crossmodal effect underlies the preference of the visual cortex for the pre-learning condition. In the auditory cortex, a mirror pattern appeared with visual attenuation before learning and not afterwards. This release of attenuation explains the preference in the auditory cortex for the post-learning. In the Plus experiment, both crossmodal effects were apparent.

informative stimuli, whereas the visual stimuli remained unchanged. These shifts were detected in associative areas and were mostly lateralized to the left hemisphere, in areas previously reported to exhibit multisensory responses (Jones and Powell 1970; Calvert 2001; Fairhall and Macaluso 2009;

Beauchamp et al. 2010; Noppeney et al. 2010). Clusters were located in the left prefrontal cortex, the left IFS, the IFG and the Middle Frontal Gyrus (MFG), the bilateral anterior Supramarginal Gyrus (aSMG), and the left IPS. A close examination of the beta values within these areas revealed a shift in

sensory preference. With the exception of the left IPS, all the regions demonstrated a significant ($P < 0.05$, FDR corrected) response to visual stimuli but not to auditory stimuli before learning, and to auditory stimuli but not to visual after

learning. The left IPS exhibited a significant response to both visual and auditory stimuli before and after learning, but with changes in the magnitude of the effect (lower mean response to pre-auditory and post-visual stimuli; Fig. 4B).

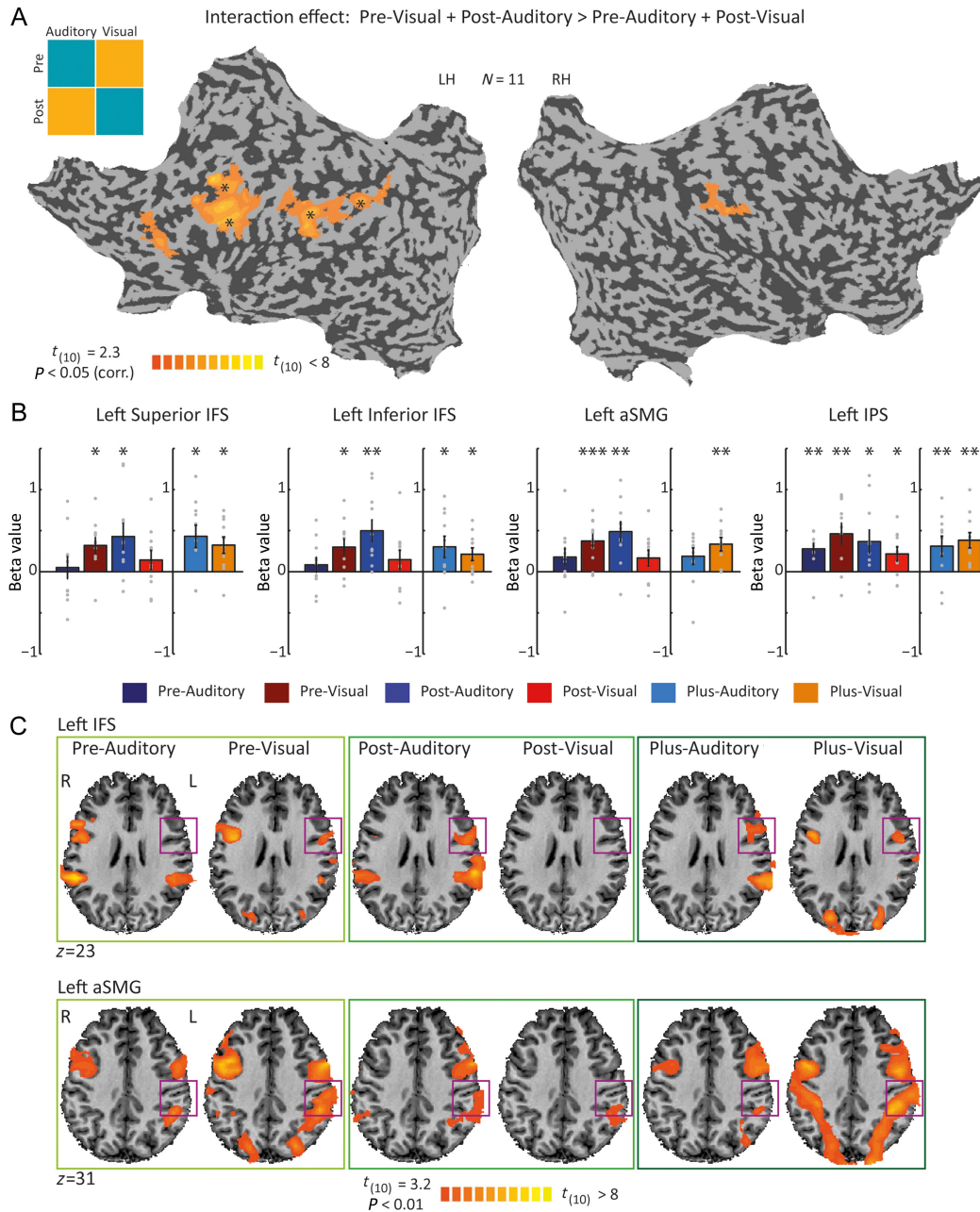


Figure 4. Dynamic shift in sensory responses outside sensory areas. (A) Statistical parametric map of the interaction effect (Modality \times Learning) revealed by a two-way ANOVA analysis, Pre-Visual + Post-Auditory versus Post-Visual + Pre-Auditory ($P < 0.05$, corr.), is presented on a flattened cortical reconstruction of one of the subjects. The analysis was carried out within areas that were responsive to either vision or auditory, before or after learning SSA. Positive responses represent areas which were more responsive to visual stimuli than auditory stimuli before learning, but more responsive to auditory stimuli than visual stimuli after learning. In the pre-learning condition, only visual input is informative, while after learning SSA auditory soundscapes input can be deciphered to reveal shape information. Cortical areas that shifted their sensory preferences from one informative sensory input to the other after learning include the left IFS, the left aSMG, the left and right IPS, and the left MFG. (B) Beta values were sampled from the clusters depicted in the statistical parametric map, and presented for the clusters in which both Pre-Visual and Post-Auditory were statistically significant ($*P < 0.05$, $**P < 0.005$, $***P < 0.0005$, corr.). The location of these clusters is marked with black asterisks on the flat cortical reconstructions. Left IFS clusters and left IPS also showed significant responses to both auditory and visual stimuli in the "Plus" experiment, in which information from both modalities had to be used to perform the task. (C) Auditory and visual responses outside sensory areas (compared with baseline, $P < 0.01$, uncorrected). Left IFS ($z = 23$, top row) and left aSMG ($z = 31$, bottom row) were found to change their sensory preference across experiments (marked by a rectangle). Before learning SSA both were significantly responsive to visual stimuli, but not significantly to auditory stimuli. After learning they became significantly responsive to auditory stimuli, but not to visual stimuli. In the Plus detection experiment, the left IFS was significantly responsive to both auditory and visual stimuli, whereas the left aSMG preferred visual stimuli. Both areas could access both visual and auditory inputs, but changed their sensory preference according to the information, novelty, and task relevance of the sensory input.

This shift was further demonstrated by the auditory and visual response maps. The statistical parametric maps describe the significant response to either the auditory or visual condition (significant difference from zero, see Methods; Fig. 4C). The left IFS was significantly responsive to visual but not to auditory stimuli before learning the SSA (“Pre”) and was significantly responsive to auditory but not to visual stimuli after learning the SSA (“Post”). This pattern was similar in other areas, thus showing a shift in sensory preference, as illustrated in the beta value plots (Fig. 4B). In these areas, sensory preference adhered to the changes in information, novelty, and task relevance of the sensory inputs of the experimental design. Specifically, before SSA learning, visual input was informative and auditory input was undecipherable, and these areas were visually responsive. After learning the SSA, the auditory soundscape contained novel shape information, and sensory preference shifted to auditory responses.

Sensory responses in these associative areas were also examined during the “Plus” experiment. During the audiovisual integration task, the responses to auditory and visual stimuli could not be distinguished. With the exception of the aSMG (which showed a marginally significant response to auditory stimuli and a significant visual response), all associative areas tested showed significant responses to both auditory and visual stimuli, with similar magnitude (Fig. 4B). To further confirm the associative area responses during the “Plus” experiment, “Plus-Auditory” and “Plus-Visual” conditions were statistically assessed (Fig. 4C and Table 2). In the left IFG, both auditory and visual stimuli elicited significant responses ($P < 0.01$, Fig. 4C, top row). In the aSMG, auditory responses were below threshold, whereas visual responses exceeded threshold ($P < 0.01$, Fig. 4C, bottom row). These findings therefore suggest that when both visual and auditory inputs were task relevant, associative areas were similarly responsive to both visual and auditory stimuli.

Auditory and Visual Overlap

Finally, consistent audiovisual convergence was examined by using a probabilistic map overlapping all auditory and visual conditions from the 3 experiments. This analysis was carried out to detect areas that did not change their multisensory preference, that is, areas that responded similarly to auditory and visual inputs, and thus could not be detected by any of the ANOVA measures detailed above (modality, learning, or interaction effect). It could also reveal areas that did indeed demonstrate changes in sensory response profiles, but were significantly responsive to both sensory inputs. Full audiovisual convergence throughout the experiments was found in the right Middle Temporal Gyrus (MTG) and left IPS (Fig. 5). These areas showed consistent audiovisual convergence regardless of the experimental conditions in which the stimuli were delivered, with positive auditory and visual responses whenever these were present. The right MTG responded similarly to auditory and visual stimuli, regardless of their experimental context (Fig. 5B, top row), showing clear audiovisual overlap, in line with a hierarchical, bottom-up view of multisensory processing. The left IPS also showed an overlap of auditory and visual responses (Fig. 5B, bottom row). However, this area also demonstrated modulation of sensory preference on top of these consistent audiovisual convergence responses, as emerged from the ANOVA interaction effect (Fig. 4). While

showing both significant auditory and visual responses throughout the experiments, the relations between these changed, in that visual responses were higher than auditory responses before learning, lower after learning, and equal during the audiovisual detection task. This consistent overlap supports a bottom-up view of multisensory perception, in that sensory inputs are processed in sensory areas and then converge in associative areas (Beauchamp et al. 2004, 2010; Calvert and Thesen 2004; van Atteveldt et al. 2004; Hertz and Amedi 2010).

Discussion

We examined the ways in which task relevance and information conveyed by sensory inputs affect their processing in sensory and associative areas. We hypothesized that the context in which sensory inputs are delivered can account for the variability in crossmodal effects in sensory cortices and in sensory responses in associative areas found in the literature, including our own lab’s previous findings, and demonstrate how multisensory processing can be both a highly dynamic system and a deterministic and hierarchical one.

To this end, 3 experimental conditions were carried out, which were similar or identical in terms of stimuli, but where the information, novelty, and task relevance of the sensory inputs were manipulated. Subjects were presented with visual images and auditory soundscapes before and after SSA learning, and during an audiovisual integration task in which a specific combination of soundscape and image (resulting with a + sign, see Methods) had to be detected. The learning period was very short—subjects were scanned before and after 1 h of a training session on the SSA, such that no long-term plasticity could be manifested. Auditory and visual responses were examined in associative and sensory areas. Sensory areas did not change their sensory responses; for example, the auditory cortex showed strong auditory responses and the visual cortex showed strong visual responses throughout the experiments (Fig. 2). However, crossmodal effects in sensory areas changed their pattern and direction rapidly according to the experimental conditions (Fig. 3). A number of associative cortical regions, including the left IFG, the left aSMG, and the IPS demonstrated a shift in their sensory responses when the sensory input context changed (Fig. 4). In contrast, consistent, multisensory overlap was detected in the right MTG and the left IPS (Fig. 5). These results are suggestive of the Janus-like nature of multisensory processing—a stable, bottom-up system, which is not context-dependent, in line with the classical view, and dynamic context-dependent elements, which modulate the stable system. It also demonstrates that the experimental context in which sensory inputs are delivered plays an important role in sensory processing, even within sensory areas, and can explain the variability in sensory responses and multisensory interactions described in the literature. Finally, our results do not exclude the possibility of long-term plasticity of the system, following prolonged learning, or in cases of radical changes in the sensory input as in sensory deprivation (Bavelier et al. 2006, 2012; Reich et al. 2011; Striem-Amit et al. 2012a).

The use of a visual-to-auditory SSA was crucial to obtaining these results, since it enabled the manipulation of information conveyed by the auditory input while keeping the stimuli intact after only 1 h of learning. The effect of learning a new

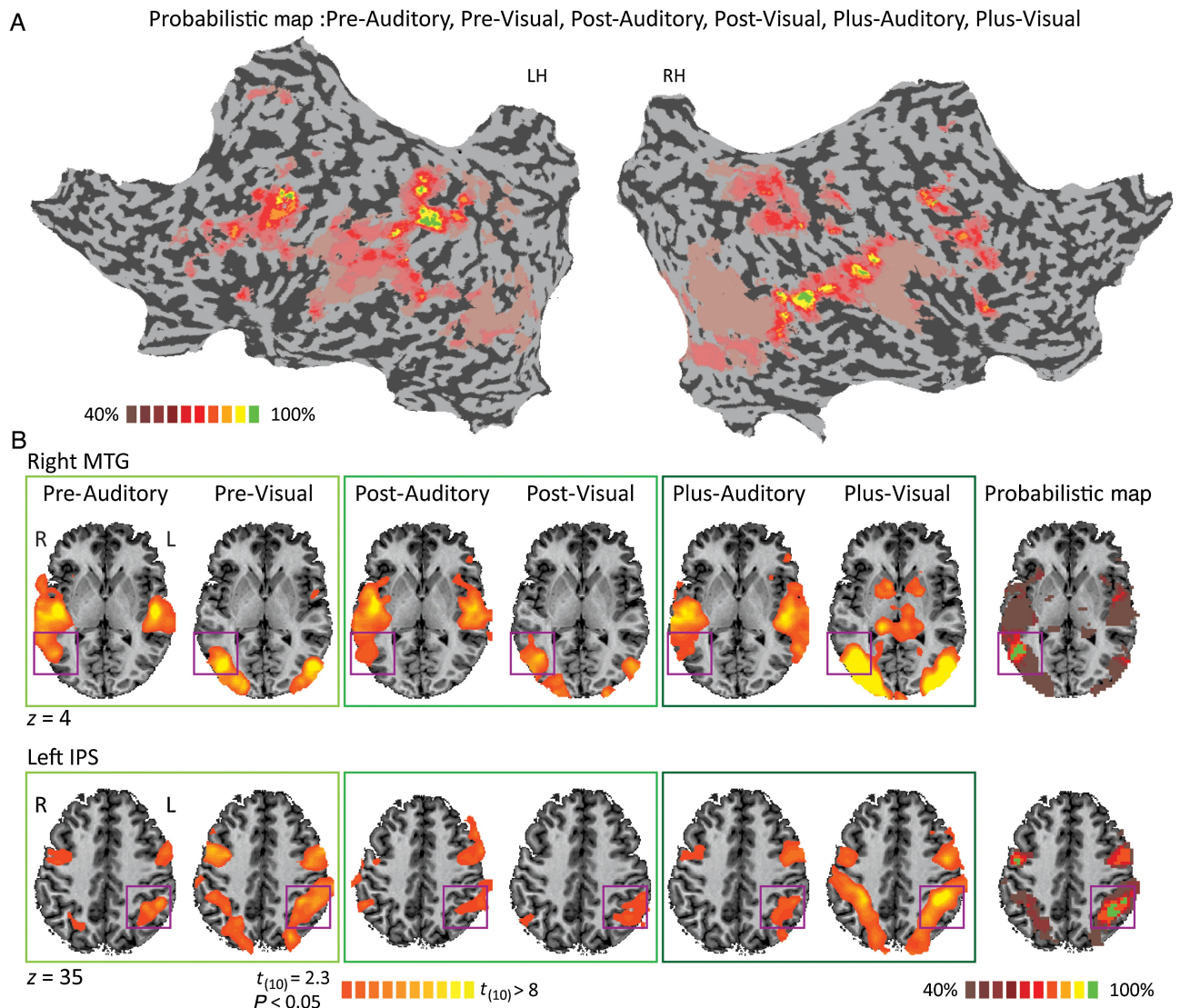


Figure 5. Consistent audiovisual responses outside sensory areas. (A) A probabilistic overlap map, created from auditory and visual responses in the Pre, Post, and Plus experiments, presented on a flattened cortical reconstruction of one of the subjects. 100% overlap, marked in green, was found in the right MTG, left IPS, and left MFG, and means that these areas showed significant responses to auditory and visual stimuli in all 3 experimental conditions. (B) Auditory and visual response maps from all the experiments are presented, demonstrating consistent auditory- and visual-positive responses in the right MTG ($z = 4$, top row) and the left IPS ($z = 35$, bottom row) (compared with baseline, $P < 0.05$, uncorrected). These areas were responsive to auditory and visual stimuli regardless of the experimental context, task relevance, or the information they conveyed.

audiovisual association has been studied before (Naumer et al. 2009). The results have indicated changes in the audiovisual integration of learned audiovisual associations. However, learning an arbitrary audiovisual association is different from learning a visual-to-auditory SSA. First, subjects were able to extract visual information from soundscapes after very short learning periods (Kim and Zatorre 2008; Striem-Amit et al. 2012b), whereas audiovisual associations rely either on life-long experience (for example, in the case of letters and phonemes) or on long periods of training (Naumer et al. 2009). Another line of studies demonstrated this point by testing the recognition of images which were initially viewed without accompanying sound, with an ecological accompanying sound (an image of a bell and a “dong” sound), and images accompanied by arbitrary meaningless sounds, without any training (Murray et al. 2004, 2005; Thelen et al. 2012). These short audiovisual presentations affected image recognition

when viewed later without their accompanying sounds, in that ecological sounds facilitated recognition, but meaningless sounds impaired recognition, showing that even a brief audiovisual presentation can affect behavior, but that this behavior is tightly linked to the ecological relation between auditory and visual stimuli. SSAs therefore provide a unique possibility to examine auditory stimuli, which were meaningless before becoming informative. SSAs were also shown to be processed differently from associations, as listening to sounds that were learned to be associated with visual images failed to activate the visual cortex, whereas listening to SSA produced soundscapes activated the visual cortex (Amedi et al. 2007). The ability of SSAs to activate the visual cortex in a very specific manner was reported in other studies as well, using either visual to auditory transformations such as the one used here, or tactile sensory substitution devices that transform visual information to grid tactile actuators. These include activation

of area MT for motion (Ptito et al. 2009; Matteau et al. 2010), area lateral occipital complex for shape (Amedi et al. 2007; Kim and Zatorre 2011; Striem-Amit et al. 2012b), and area MOG for localization (Renier et al. 2010; Collignon et al. 2011; Striem-Amit et al. 2012b).

A number of confounds and caveats should be kept in mind when interpreting the results. While we cannot claim that auditory soundscapes were perceived as images by our subjects, judging by their performance on the audiovisual integration task we can assume that they were able to extract some very simple shape information from the soundscapes after 1 h of learning. We also cannot rule out the possibility and role of visual imagery during perception of auditory soundscapes after learning SSA. It may be that the shift in sensory responses in associative areas was driven by the visual imagery of auditory soundscapes. However, previous studies have demonstrated that visual imagery can result in the deactivation of auditory cortex and positive responses in the visual cortex, the opposite pattern to the one described here (Amedi et al. 2005). In addition, visual imagery had to compete with the visual stimuli that were still present on the screen. In this case, one may expect to find positive auditory responses in visual areas such as LOC (Amedi et al. 2007), which was not seen here. In any case, our results show that, after learning, associative areas were more responsive to information conveyed by auditory input than that conveyed by visual input. Finally, a confound of every learning design, including the one used here, is that post-learning sessions always come after the pre-learning, so fatigue and the simple passage of time may affect the results. However, fatigue or boredom is likely to cause overall changes in sensory processing and stimulus-locked activity throughout the brain, and affect both auditory and visual processing in a similar manner. Our results, however, indicate a reverse theme for auditory and visual responses, as one increases and the other decreases. Whereas fatigue cannot be explicitly ruled out in this design, the resulting effect seems to have followed the changes in information and the novelty of sensory inputs rather than decline regardless of sensory input. Taking all these considerations into account does not change our basic observations, and allows us to draw some conclusions regarding dynamic and consistent multisensory processing.

The results showed that while the sensory cortices did not change their sensory preference, their response was dramatically affected by the competing stimuli. Evidence of crossmodal effects in primary sensory areas have accumulated in recent years and have been reported under a variety of experimental conditions in rats, primates, and humans, with different and sometime contradictory effects (Schroeder and Foxe 2005; Ghazanfar and Schroeder 2006; Kayser 2010). Our results, demonstrating changes in the direction of crossmodal responses according to the experimental condition, support the emerging notion that crossmodal effects in sensory areas are context and history dependent (in contrast to their constant main sensory preference, in line with their role in the classic/traditional, hierarchical scheme of sensory processing). Crossmodal enhancement has been seen in experiments where one sensory input carries information associated with another sensory input; for example, watching touch or lip reading (Calvert et al. 1997; Blakemore et al. 2005) or when this input conveys a special ecological, or task-specific meaning such as pup odor enhancing responses to pup's cries in the auditory cortex of mouse mothers (Cohen et al. 2011), or when inputs are in

spatial or temporal congruency (Lakatos et al. 2007). However, when sensory inputs compete with one another, or are incongruent, crossmodal attenuations or deactivations are apparent. For example, during visual imagery, the auditory cortex is inhibited, and when trying to understand speech in a loud environment the visual cortex is inhibited (Kuchinsky et al. 2012). Deactivation of primary sensory areas was also apparent during a nonsensory task, in verbal memory (Azulay et al. 2009). Crossmodal attenuation is commonly described in terms of attention: when a task demands attention to a specific modality, crossmodal inhibition of the irrelevant input is seen (Gazzaley and D'Esposito 2007; Talsma et al. 2010). It is important to note that while the underlying mechanisms of the negative BOLD signal are still being studied (Shmuel et al. 2006; Schridde et al. 2008; Azulay et al. 2009), the above studies and others show that these deactivations have functional implications (Kastrup et al. 2008; Diedrichsen et al. 2013; Zeharia et al. 2012). Here, however, crossmodal attenuations did not represent negative BOLD *per se*, but rather manifested in a decrease of positive responses in the presence of competing stimuli, as both auditory and visual stimuli were presented in the same experiment. When visual stimuli were attenuated by auditory stimuli (after SSA learning), they still showed a positive response to visual stimuli, as was found in the modality effect (Fig. 2), but were attenuated when auditory stimuli were presented, resulting in negative beta values for auditory stimuli, along with positive beta values for visual stimuli (see Fig. 3B, and compare Figs 2B and 3C). These crossmodal attenuations changed direction in a context-dependent manner, consistent with the changes in novelty, task relevance, and information carried by each sensory input. Crossmodal attenuations may serve as a mechanism that filters sensory inputs, either noisy, ambiguous or task irrelevant, thus allowing for more efficient processing higher along the hierarchy (Corbetta and Shulman 2002; Gazzaley and D'Esposito 2007; Bressler et al. 2008; Magosso et al. 2010; Werner and Noppeney 2010). Our results suggest that this mechanism is dynamic and can change its sensory direction and effect quickly in response to changes in task context and the relevance of the sensory inputs. Similarly, although not found here, crossmodal effects may serve as a mechanism for enhancing the perception of one sensory modality via crossmodal enhancements.

Associative areas demonstrated dramatic shifts in sensory preference, in that they changed the strength of the response to sensory inputs throughout the experiments. These areas were mainly left lateralized, including the prefrontal, IFS, aSMG, and IPS. This left lateralized network overlapped with other known left lateralized networks involved in object detection (Werner and Noppeney 2010; Lee and Noppeney 2011), SSA learning (Amedi et al. 2007), attention (Corbetta and Shulman 2002), and language (Hickok and Poeppel 2000). The left IPS, IFS, and lateral occipital cortex (LO) were shown to be activated during object detection, either when object images were visually presented, when they were touched, or via auditory SSA soundscapes (Amedi et al. 2007). However, these multisensory responses were found when the information the sensory inputs conveyed was kept constant—they all carried shape information—and when sensory inputs were delivered one at a time, and never competed with each other. In this respect, reports elsewhere of multisensory responses in the left IFS cannot be differentiated from the multisensory responses in STS (van Atteveldt et al. 2004; Beauchamp et al.

2008; Stevenson et al. 2009). In these studies, the STS was found to be activated by both auditory and visual stimuli when they were delivered in different blocks (audiovisual convergence) and when delivered together (sometimes exhibiting a superadditivity response). However, it is only when the experimental context was manipulated and when the stimuli remained the same and were delivered together that changes in the nature of the multisensory response were found: The right MTG (adjacent to the STS) showed a noncontext-dependent response, whereas the left IFS changed its sensory preference in a context-dependent manner. This differentiation between the temporal multisensory regions and the parietal and frontal multisensory areas may indicate different roles in multisensory processing, and its interaction with other cognitive systems such as attention (Talsma et al. 2010). The MTG may be a deterministic, bottom-up convergence area, in which auditory and visual inputs converge regardless of task, information, or novelty. It may be sensitive to the physical aspects of sensory inputs, such as temporal alignment or noise. The parietal and frontal areas may be directed toward specific audiovisual combination, and be affected by the information, task relevance, and/or other contextual cues.

The left aSMG and the left IFS also play an important role in audiovisual speech perception (Hickok and Poeppel 2000). The left IFS is classically present in speech perception models (Broca's area) (Hickok and Poeppel 2000; Calvert and Campbell 2003; Lee and Noppeney 2011). The left aSMG was also shown to be important in the detection of audiovisual speech cues (Wernike's area) (Calvert and Campbell 2003; Bernstein et al. 2008). Furthermore, both areas play an important role in audiovisual speech perception (Callan, Tajima, et al. 2003; Lee and Noppeney 2011; Tkach et al. 2011). Our results have something in common with these language and shape detection studies. The SSA soundscapes and visual images used in this study convey shape information, and integration of audiovisual shapes was explicitly demanded in our active plus detection experiment. SSA is also similar to language in that abstract information is extracted from auditory input based on a set of rules and principles; namely, linguistic/semantic information or visual information. The similarity in cortical areas between these language areas, object detection areas, and our results may suggest that these areas play a general role in extracting abstract meaning from sensory information, and can change their sensory preference based on saliency, information, or ecological context. This notion is in line with the neural recycling hypothesis (Dehaene and Cohen 2007). According to this hypothesis, a novel cultural object encroaches onto a pre-existing brain system, resulting in an extension of existing cortical networks to address the novel functionality. This hypothesis was used to explain the cortical network dedicated to reading and show how novel cultural objects rely on older brain circuits used for objects and face recognition (Dehaene et al. 2005). Such recycling of multisensory object detection networks as was reported here, which emphasize the extraction of abstract information from sensory inputs, might take place in the formation of cortical networks dedicated to language processing.

A line of studies support the notion of dynamic weighting of sensory inputs according to the input reliability (Ernst and Banks 2002; Shams 2012). Behavioral measures showed that subjects integrate multisensory inputs in a statistically optimal manner, assigning lower weights to noisy inputs (Ernst and Banks 2002; Sheppard et al. 2013). Recently, dynamic

weighting was identified in the dorsal medial temporal area of the monkey (Fetsch et al. 2009, 2012). Neuroimaging studies in humans have revealed dynamic weighting of sensory inputs in the IPS, which showed a change in preference for visual and tactile inputs as their reliability was manipulated (Beauchamp et al. 2010). In another study, the role of the left IFS in decisions about audiovisual object categorization was examined by manipulating the saliency of sensory inputs. It changed its weighting of sensory inputs dynamically based on the sensory input saliency: when auditory input was noisy, the left IFS preferred visual input and vice versa (Adam and Noppeney 2010; Noppeney et al. 2010; Lee and Noppeney 2011). Our results support the notion of dynamic weighting of sensory input in the IFS and IPS associative areas. Our study extends previous results in that the stimuli remained the same, and information conveyed by sensory inputs was not manipulated by adding noise, but by SSA learning. It is therefore not only the physical aspects of the stimuli that impact the dynamic weighting process, but also the information they convey.

Overall, these results show that changes in sensory responses in associative areas were accompanied with crossmodal attenuations of sensory cortices. A number of explanations can link these observations with previous results. The first is that dynamic weighting is determined by the responses of the sensory cortices. Adding noise to a sensory input may alter the signal delivered by the sensory cortex. Similarly, attenuation of the sensory cortex may alter that signal. In both cases, associative areas may prefer the more informative, unaltered signal (if there is one). Crossmodal attenuations can be a filter or a gating mechanism controlled by attentive processes, resulting in the same effect as adding noise to a sensory input. Another possibility is that a common source drives both processes, for example, the crossmodal attenuation pattern and the weighting of sensory inputs. These may facilitate each other, leading to more efficient multisensory perception. Our speculation is that the content of sensory inputs plays an important role in determining crossmodal attenuations and sensory weighting, and therefore higher cognitive control mechanisms are crucial to these processes.

To conclude, in this study, we examined whether the experimental context and information conveyed by sensory inputs could account for several key questions in sensory processing and integration. Our results suggest that multisensory perception involves 2 components. One is a deterministic, noncontext-dependent bottom-up structure, including consistent sensory preferences in primary sensory areas, and multisensory convergence in associative areas such as the MTG/STS (between the occipital visual and temporal auditory cortices) for audiovisual integration. The second is highly dynamic and context-dependent, and includes crossmodal modulation of sensory areas, and a dynamic selection of sensory inputs in associative areas, which are mostly left lateralized. We were able to link these 2 phenomena, which to date have only been reported separately and independently of each other. Recently, some organizing principles of multisensory perception have been suggested, some of which highlight the potential of the entire brain for multisensory responses (Ghazanfar and Schroeder 2006; Kayser 2010), by using different principles to determine the sensory preference and responses; for example, relying on the information conveyed by a stimulus instead of its modality (Pascual-Leone and Hamilton 2001), or using attention as an explaining mechanism (Talsma et al. 2010). We suggest that, in multisensory perception, there are both stable and dynamic

processes, allowing stability and a fast, dynamic tuning of the system when required; both can take place together, which may lead to seemingly contradictory evidence. Whereas long-term training or sensory deprivation have been shown to induce long-term plasticity of sensory processing mechanisms, here we showed that even in the very short term, multisensory processing is a plastic process. This dynamic nature allows for rapid adaptation of sensory processing to a noisy and changing world, and may also serve as a basis for the evolution of niche culture abilities such as language.

Funding

This work was supported by a career development award from the International Human Frontier Science Program Organization (HFSP), the Israel Science Foundation (grant no. 1530/08), a European Union Marie Curie International Reintegration Grant (MIRG-CT-2007-205357), a James S. McDonnell Foundation scholar award (grant no. 220020284), the Edmond and Lily Safra Center for Brain Sciences Vision center grant, and the European Research Council grant (grant no. 310809) (to A.A.).

Notes

U.H. wishes to acknowledge the Charitable Gatsby Foundation for its support. *Conflict of Interest*: None declared.

References

- Adam R, Noppeney U. 2010. Prior auditory information shapes visual category-selectivity in ventral occipito-temporal cortex. *Neuroimage*. 52:1592–1602.
- Adrian ED. 1949. The Sherrington Lectures. I. Sensory integration. Liverpool (UK): University of Liverpool Press.
- Amedi A, Malach R, Pascual-Leone A. 2005. Negative BOLD differentiates visual imagery and perception. *Neuron*. 48:859–872.
- Amedi A, Stern WM, Camprodon JA, Bertolino A, Merabet L, Rotman S, Hemond C, Meijer P, Pascual-Leone A, Bempohl F. 2007. Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat Neurosci*. 10:687–689.
- Azulay H, Striem E, Amedi A. 2009. Negative BOLD in sensory cortices during verbal memory: a component in generating internal representations? *Brain Topogr*. 21:221–231.
- Bavelier D, Dye MWG, Hauser PC. 2006. Do deaf individuals see better? *Trends Cogn Sci*. 10:512–518.
- Bavelier D, Green CS, Pouget A, Schrater P. 2012. Brain plasticity through the life span: learning to learn and action video games. *Annu Rev Neurosci*. 35:391–416.
- Bavelier D, Neville HJ. 2002. Cross-modal plasticity: where and how? *Nat Rev Neurosci*. 3:443–452.
- Beauchamp MS, Lee KE, Argall BD, Martin A. 2004. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*. 41:809–823.
- Beauchamp MS, Pasalar S, Ro T. 2010. Neural substrates of reliability-weighted visual-tactile multisensory integration. *Front Syst Neurosci*. 4:25.
- Beauchamp MS, Yasar NE, Frye RE, Ro T. 2008. Touch, sound and vision in human superior temporal sulcus. *Neuroimage*. 41:1011–1020.
- Benevento LA, Fallon J, Davis BJ, Rezak M. 1977. Auditory–visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Exp Neurol*. 57:849–872.
- Bernstein LE, Lu Z-L, Jiang J. 2008. Quantified acoustic-optical speech signal incongruity identifies cortical sites of audiovisual speech processing. *Brain Res*. 1242:172–184.
- Blakemore S-J, Bristow D, Bird G, Frith C, Ward J. 2005. Somatosensory activations during the observation of touch and a case of vision-touch synaesthesia. *Brain*. 128:1571–1583.
- Boynton GM, Engel SA, Glover GH, Heeger DJ. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci*. 16:4207–4221.
- Bressler SL, Tang W, Sylvester CM, Shulman GL, Corbetta M. 2008. Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *J Neurosci*. 28:10056–10061.
- Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E. 2003. Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport*. 14:2213–2218.
- Callan DE, Tajima K, Callan AM, Kubo R, Masaki S, Akahane-Yamada R. 2003. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage*. 19:113–124.
- Calvert GA. 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex*. 11:1110–1123.
- Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS. 1999. Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport*. 10:2619–2623.
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS. 1997. Activation of auditory cortex during silent lipreading. *Science*. 276:593–596.
- Calvert GA, Campbell R. 2003. Reading speech from still and moving faces: the neural substrates of visible speech. *J Cogn Neurosci*. 15:57–70.
- Calvert GA, Campbell R, Brammer MJ. 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol*. 10:649–657.
- Calvert GA, Thesen T. 2004. Multisensory integration: methodological approaches and emerging principles in the human brain. *J Physiol Paris*. 98:191–205.
- Cohen L, Rothschild G, Mizrahi A. 2011. Multisensory integration of natural odors and sounds in the auditory cortex. *Neuron*. 72:357–369.
- Cohen LG, Celnik P, Pascual-Leone A, Corwell B, Falz L, Dambrosia J, Honda M, Sadato N, Gerloff C, Catalá MD et al. 1997. Functional relevance of cross-modal plasticity in blind humans. *Nature*. 389:180–183.
- Collignon O, Vandewalle G, Voss P, Albouy G, Charbonneau G, Lassonde M, Lepore F. 2011. Functional specialization for auditory-spatial processing in the occipital cortex of congenitally blind humans. *Proc Natl Acad Sci USA*. 108:4435–4440.
- Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*. 3:201–215.
- Dehaene S, Cohen L. 2007. Cultural recycling of cortical maps. *Neuron*. 56:384–398.
- Dehaene S, Cohen L, Sigman M, Vinckier F. 2005. The neural code for written words: a proposal. *Trends Cogn Sci*. 9:335–341.
- Diedrichsen J, Wiestler T, Krakauer JW. 2013. Two distinct ipsilateral cortical representations for individuated finger movements. *Cereb Cortex*. 23:1362–1377.
- Engel SA, Glover GH, Wandell BA. 1997. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex*. 7:181–192.
- Ernst MO, Banks MS. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*. 415:429–433.
- Fairhall SL, Macaluso E. 2009. Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *Eur J Neurosci*. 29:1247–1257.
- Felleman DJ, Van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*. 1:1–47.
- Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE. 2012. Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci*. 15:146–154.
- Fetsch CR, Turner AH, DeAngelis GC, Angelaki DE. 2009. Dynamic reweighting of visual and vestibular cues during self-motion perception. *J Neurosci*. 29:15601–15612.
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med*. 33:636–647.

- Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE. 2000. Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res Cogn Brain Res*. 10:77–83.
- Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D, Ritter W, Murray MM. 2002. Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J Neurophysiol*. 88:540–543.
- Friston KJ, Holmes AP, Price CJ, Büchel C, Worsley KJ. 1999. Multisubject fMRI studies and conjunction analyses. *Neuroimage*. 10:385–396.
- Friston KJ, Holmes AP, Worsley KJK, Poline JJ-BJ-P, Frith CD, Frackowiak RSJ. 1994. Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp*. 2:189–210.
- Friston KJ, Jezzard P, Turner R. 1994. Analysis of functional MRI time-series. *Hum Brain Mapp*. 12:153–171.
- Fu K-MG, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garaghty PE, Schroeder CE. 2003. Auditory cortical neurons respond to somatosensory stimulation. *J Neurosci*. 23:7510–7515.
- Gazzaley A, D'Esposito M. 2007. Top-down modulation and normal aging. *Ann N Y Acad Sci*. 1097:67–83.
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK. 2005. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci*. 25:5004–5012.
- Ghazanfar AA, Schroeder CE. 2006. Is neocortex essentially multisensory? *Trends Cogn Sci*. 10:278–285.
- Hertz U, Amedi A. 2010. Disentangling unisensory and multisensory components in audiovisual integration using a novel multi-frequency fMRI spectral analysis. *Neuroimage*. 52:617–632.
- Hickok G, Poeppel D. 2000. Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci*. 4:131–138.
- Jones EG, Powell TPS. 1970. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain*. 93:793–820.
- Kastrup A, Baudewig J, Schnaudigel S, Huonker R, Becker L, Sohns JM, Dechent P, Klingner C, Witte OW. 2008. Behavioral correlates of negative BOLD signal changes in the primary somatosensory cortex. *Neuroimage*. 41:1364–1371.
- Kayser C. 2010. The multisensory nature of unisensory cortices: a puzzle continued. *Neuron*. 67:178–180.
- Kayser C, Logothetis NK. 2007. Do early sensory cortices integrate cross-modal information? *Brain Struct Funct*. 212:121–132.
- Kayser C, Petkov CI, Logothetis NK. 2008. Visual modulation of neurons in auditory cortex. *Cereb Cortex*. 18:1560–1574.
- Kim J-K, Zatorre RJ. 2008. Generalized learning of visual-to-auditory substitution in sighted individuals. *Brain Res*. 1242:263–275.
- Kim J-K, Zatorre RJ. 2011. Tactile-auditory shape learning engages the lateral occipital complex. *J Neurosci*. 31:7848–7856.
- Kuchinsky SE, Vaden KI, Keren NI, Harris KC, Ahlstrom JB, Dubno JR, Eckert MA. 2012. Word intelligibility and age predict visual cortex activity during word listening. *Cereb Cortex*. 22:1360–1371.
- Lakatos P, Chen C, O'Connell MN, Mills A, Schroeder CE, Connell MNO. 2007. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*. 53:279–292.
- Laurienti PJ, Burdette JH, Wallace MT, Yen Y, Field AS, Stein BE. 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. *J Cogn Neurosci*. 14:420–429.
- Lee H, Noppeney U. 2011. Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *J Neurosci*. 31:11338–11350.
- Magosso E, Serino A, di Pellegrino G, Ursino M. 2010. Crossmodal links between vision and touch in spatial attention: a computational modelling study. *Comput Intell Neurosci*. 2010:304941.
- Matteau I, Kupers R, Ricciardi E, Pietrini P, Ptito M. 2010. Beyond visual, aural and haptic movement perception: hMT+ is activated by electro-tactile motion stimulation of the tongue in sighted and in congenitally blind individuals. *Brain Res Bull*. 82:264–270.
- Meijer PB. 1992. An experimental system for auditory image representations. *IEEE Trans Biomed Eng*. 39:112–121.
- Merabet LB, Rizzo JF, Amedi A, Somers DC, Pascual-Leone A. 2005. What blindness can tell us about seeing again: merging neuroplasticity and neuroprostheses. *Nat Rev Neurosci*. 6:71–77.
- Meredith MA, Stein BE. 1986. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol*. 56:640–662.
- Mozolic JL, Joyner D, Hugenschmidt CE, Peiffer AM, Kraft RA, Maldjian JA, Laurienti PJ. 2008. Cross-modal deactivations during modality-specific selective attention. *BMC Neurol*. 8:35.
- Murray MM, Foxe JJ, Wylie GR. 2005. The brain uses single-trial multisensory memories to discriminate without awareness. *Neuroimage*. 27:473–478.
- Murray MM, Michel CM, Grave de Peralta R, Ortigue S, Brunet D, Gonzalez Andino S, Schneider A. 2004. Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage*. 21:125–135.
- Naumer MJ, Doehrmann O, Müller NG, Muckli L, Kaiser J, Hein G. 2009. Cortical plasticity of audio-visual object representations. *Cereb Cortex*. 19:1641–1653.
- Noppeney U, Ostwald D, Werner S. 2010. Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *J Neurosci*. 30:7434–7446.
- Pascual-Leone A, Amedi A, Fregni F, Merabet LB. 2005. The plastic human brain cortex. *Annu Rev Neurosci*. 28:377–401.
- Pascual-Leone A, Hamilton R. 2001. The metamodal organization of the brain. *Prog Brain Res*. 134:427–445.
- Ptito M, Matteau I, Gjedde A, Kupers R. 2009. Recruitment of the middle temporal area by tactile motion in congenital blindness. *Neuroreport*. 20:543–547.
- Reich L, Szved M, Cohen L, Curie M, Amedi A. 2011. A ventral visual stream reading center independent of visual experience. *Curr Biol*. 21:1–6.
- Renier LA, Anurova I, De Volder AG, Carlson S, VanMeter J, Rauschecker JP. 2010. Preserved functional specialization for spatial processing in the middle occipital gyrus of the early blind. *Neuron*. 68:138–148.
- Schridde U, Khubchandani M, Motelow JE, Sanganahalli BG, Hyder F, Blumenfeld H. 2008. Negative BOLD with large increases in neuronal activity. *Cereb Cortex*. 18:1814–1827.
- Schroeder CE, Foxe J. 2005. Multisensory contributions to low-level, “unisensory” processing. *Curr Opin Neurobiol*. 15:454–458.
- Sekuler R, Sekuler AB, Lau R. 1997. Sound alters visual motion perception. *Nature*. 385:308.
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB. 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*. 268:889–893.
- Shams L. 2012. Early integration and Bayesian causal inference in multisensory perception. In: Murray MM, Wallace MT, editors. *The neural bases of multisensory processes*. Boca Raton (FL): CRC Press.
- Shams L, Kamitani Y, Shimojo S. 2000. Illusions. What you see is what you hear. *Nature*. 408:788.
- Sheppard JP, Raposo D, Churchland AK. 2013. Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans. *J Vis*. 13:1–19.
- Shmuel A, Augath M, Oeltermann A, Logothetis NK. 2006. Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nat Neurosci*. 9:569–577.
- Stevenson RA, Kim S, James TW. 2009. An additive-factors design to disambiguate neuronal and areal convergence: measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. *Exp Brain Res*. 198:183–194.
- Striem-Amit E, Cohen L, Dehaene S, Amedi A. 2012a. Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron*. 76:640–652.
- Striem-Amit E, Dakwar O, Reich L, Amedi A. 2012b. The large-scale organization of “visual” streams emerges without visual experience. *Cereb Cortex*. 22:1698–1709.
- Striem-Amit E, Hertz U, Amedi A. 2011. Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding fMRI. *PLoS ONE*. 6:e17832.
- Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG. 2010. The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci*. 14:400–410.

- Thelen A, Cappe C, Murray MM. 2012. Electrical neuroimaging of memory discrimination based on single-trial multisensory learning. *Neuroimage*. 62:1478–1488.
- Tkach JA, Chen X, Freebairn LA, Schmithorst VJ, Holland SK, Lewis BA. 2011. Neural correlates of phonological processing in speech sound disorder: a functional magnetic resonance imaging study. *Brain Lang*. 119:42–49.
- Van Atteveldt NM, Formisano E, Goebel R, Blomert L, Van Atteveldt N. 2004. Integration of letters and speech sounds in the human brain. *Neuron*. 43:271–282.
- Van Wassenhove V, Grant KW, Poeppel D, Van Wassenhove V. 2005. Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci USA*. 102:1181–1186.
- Watkins S, Shams L, Tanaka S, Haynes J-DJ, Rees G. 2006. Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage*. 31:1247–1256.
- Werner S, Noppeney U. 2010. Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *J Neurosci*. 30:2662–2675.
- Zeharia N, Hertz U, Flash T, Amedi A. 2012. Negative blood oxygenation level dependent homunculus and somatotopic information in primary motor cortex and supplementary motor area. *Proc Natl Acad Sci USA*. 109:18565–18570.
- Zeki SM. 1978. Functional specialisation in the visual cortex of the rhesus monkey. *Nature*. 274:423–428.