Funct Neurol Rehabil Ergon 2011; 1(2): 271-278 ISSN: 2156-941X

THE NEURAL NETWORK OF SENSORY-SUBSTITUTION OBJECT SHAPE RECOGNITION

Ella Striem-Amit[1], Ornella Dakwar[1], Uri Hertz[1], Peter Meijer[2], William Stern[3], Alvaro Pascual-Leone[3] and Amir Amedi[1,4*]

[1]Department of Medical Neurobiology, The Institute for Medical Research Israel-Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem 91220, Israel

[2]NXP Semiconductors, High Tech Campus 31, 5656 AE Eindhoven, The Netherlands

[3]Harvard Medical School, 330 Brookline Ave, Boston, Massachusetts 02115, USA

[4]The Edmond and Lily Safra Center for Brain Sciences (ELSC), The Hebrew University of Jerusalem, Jerusalem 91220, Israel

ABSTRACT

In sensory substitution devices (SSDs), visual information captured by an artificial receptor is delivered to the brain using non-visual sensory information. Using an auditory-to-visual SSD called "The vOICe" we previously reported that blind individuals perform successfully on object recognition tasks and are able to recruit specific ventral 'visual' structures for shape recognition using the device (i.e. through soundscapes). Comparable recruitment was also observed in sighted individuals learning to use this device. Here we directly compare a group of seven subjects who learned to perform object recognition via soundscapes and a group of seven subjects who learned arbitrary associations between sounds and object identity. We contrast these two groups' brain activity for object recognition using SSD, and for auditory object and scrambled object soundscapes. We show that the most critical structures specific for shape extraction for the purpose of object recognition are the left Pre-Central Sulcus (PCS) and the bilateral Lateral-Occipital Complex (LOC).

We also found significant activation in the occipito-parietal and posterior occipital cortex not previously observed using a smaller sample of subjects.

These results support the notion that interactions between visual structures and a network of additional areas, specifically in prefrontal cortex (PCS) might underlie the machinery which is most critical for achieving multisensory or metamodal shape recognition.

Keywords: Blindness, Cross-modal, Multisensory, Neuroimaging, Object recognition, Visual cortex.

* Correspondence: Amir Amedi, Ph.D. Dept. of Medical Neurobiology -The Institute for Medical Research Israel-Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem 91220, URL: http://brain.huji.ac.il/, E-mail: amir.amedi@ekmd.huji.ac.il

INTRODUCTION

Sensory substitution devices (SSDs) [1-5] are means of providing visual information to the blind in a non-invasive manner. Sensory substitution can be used even when other promising methods of visual restoration (i.e. surgical vision restoration, or prosthetic retinas) are unfeasible. In the case of a visual-to-auditory SSD [1], users wear a video camera linked to a computer and stereo headphones; the images are converted into “soundscapes” using a predictable algorithm, allowing them to listen to and then interpret the visual information. Remarkably, proficient users are able to differentiate the shapes of different objects, identify the actual objects, and also locate them in space. In a sense, these subjects are “seeing with sound”. Therefore, in addition to its clinical interest, sensory substitution is a valuable tool in teasing apart the influence of information modality and information content on neural processing.

Sensory substitution has already been used with some success to clarify the role of multisensory brain regions [2-5]. For example, a recent study demonstrated that during object identification, soundscapes activate the lateral-occipital tactile-visual complex (LOtv), an area which is also activated by visual and tactile object recognition [4]. As this area is known to be a region specialized in object recognition in the visual and tactile, but less so in the auditory modality [6], these findings support the notion that LOtv may be a metamodal operator [7] for shape; i.e., that it processes shape regardless of the input modality. In addition to the marked effect of shape recognition in the LOC, our previous study [4] also showed qualitatively at the single subject level that other cortical regions are also engaged in sensory-transformed object recognition such as the pre-central sulcus, inferior frontal sulcus, occipito-parietal sulcus and the posterior occipital lobe in blind individuals.

Here we directly test the role of these regions in shape recognition (as opposed to identifying objects based on arbitrary learning) using quantitative measures, in group analysis and in a contrast further controlling for the exact sensory stimulation. We directly compare a larger group of seven sighted subjects who learned to perform object recognition by extracting the object shape from the soundscapes, with a group of seven subjects who learned to identify the same soundscapes by learning arbitrary associations between sounds and object identity. Therefore, both groups heard the exact same auditory stimuli but differed in the way in which object recognition was achieved. Enlarging the group of subjects enabled us to use random effect ANOVA comparison, whose conclusions can be attributed to the entire population [8], and to study the network of regions involved in cross-modal object recognition using visual-to-auditory SSD with more statistical power.

METHODS

VISUAL-TO-AUDITORY SENSORY SUBSTITUTION

We used a visual-to-auditory sensory-substitution device (SSD) called “The vOICe” (1). The functional basis of this visuo-auditory transformation lies in spectrographic sound synthesis from any input image, which is then further perceptually enhanced through stereo panning and other techniques. Time and stereo panning constitute the horizontal axis in the sound representation of an image, tone frequency makes up the vertical axis, and loudness corresponds to pixel brightness.

PARTICIPANTS

Fourteen healthy subjects participated in the study. All subjects were right-handed as assessed by the Oldfield questionnaire, had normal hearing and normal vision (corrective lenses permitted). No subject had any known neurological or psychiatric conditions.

PROCEDURES

Sighted experts training procedures: Seven subjects were trained to interpret the shape information contained in soundscapes at the Harvard medical school. Subjects underwent training lasting 20 consecutive weekdays using a multiple choice paradigm, as well as a less structured training paradigm using a video camera linked to a laptop and stereo headphones, in which subjects were encouraged to actively explore a library of objects using The vOICe. For further details of the training procedure see [4]. All subjects achieved a minimum level of 50% success on multiple-choice testing for recognizing novel objects using "The vOICe" during training; since each question had four possible choices, this represents twice the level of success expected by chance.

Sighted association training procedures: Seven subjects were trained to arbitrarily remember and identify the soundscapes presented during the fMRI experiment. Each subject had 3-5 training sessions of 2 hours with the 8 required associations. The variable length of training represents the variable time it took the control subjects to learn the associations. The criterion for success was 100% accurate recognition on two complete sets of the 8 stimuli (16/16 associations), repeated on two consecutive days.

General Experiment Design

Each subject was scanned for 4 consecutive runs. During the scans subjects were given auditory instructions, instructing them to perform 3 tasks. (i) vOICe object recognition using SSD mapping (vOICeObj), (ii) vOICe scrambled images control (vOICeScr), (iii) auditory object recognition of 'natural' sounds made by objects (AudObj). During each condition, subjects heard two sounds of the same type (e.g. vOICe, scramble, animal/object typical sound) each repeated 3 times. Each sound lasted 2 seconds. Each pair of trials lasted a total of 12 seconds and was terminated with the auditory instruction "stop". An 8 second rest period followed, prior to the next trial. Subjects had to recognize each object and covertly name the object's identity. Stimuli were prepared representing eight objects that were recognizable both by their shape (vision transformed to vOICe) and by the sounds they make. During half of the runs, subjects were asked to identify the stimuli as '*man-made'* or '*animal',* and to respond via a two-button response box. This allowed verification that the subjects were correctly identifying objects during the scanning. To control for hand movements, subjects had to press the response buttons randomly as well during the vOICeScr condition after each cue for a stimulus.

3D recording and cortex reconstruction: Separate 3D recordings were used for surface reconstruction. High resolution 3D anatomical volumes were collected using high-resolution T1-weighted images using a 3D-turbo field echo (TFE) T1-weighted sequence (equivalent to MP-RAGE). Typical parameters were: Field of View (FOV) 23cm (RL) x 23cm (VD) x 17cm (AP); Foldover-axis: RL, data matrix: 160x160x144 zero-filled to 256 in all directions (approx 1mm isovoxel native data), TR/TE=9ms/6ms, flip angle = 8deg. Acquisition was segmented x 3 in order to enhance gray/white matter contrast. NEX = 2 in separate acquisitions. The 5 parallel imaging head coil (SENSE head) was used to reduce scan time. Reduction factor = 2.5 (total acquisition time: 5 minutes). The surface reconstruction procedure included the segmentation of the white matter of the brain of a single subject using a grow-region function. The cortical surface was then Talairach normalized (9), inflated and the obtained activation maps were superimposed onto it.

fMRI recording parameters: The BOLD fMRI measurements were performed in a whole-body 3-T, Philips scanner equipped with 22 mT/m field gradients with a slew rate of 120 T/m/s (Echospeed). The pulse sequence used was the gradient-echo echo planar imaging EPI sequence. We used 30-33 slices of 3mm thickness, with an interslice gap of 1 mm. Data in-plane matrix size was 128x128, field of view (FOV) 24cm x 24cm, time to repetition (TR) = 3000ms and time to echo (TE) = 35ms. Each experiment had 180 data points with four repetitions. The first five images (during the first baseline rest condition) were excluded from the analysis because of non-steady state magnetization.

Data analysis: Data analysis was performed using the Brain Voyager QX 1.10 software package (Brain Innovation, Maastricht, Netherlands) using standard preprocessing procedures. Functional MRI data preprocessing included head motion correction, slice scan time correction and high-pass filtering (cutoff frequency: 3 cycles/scan) using temporal smoothing in the frequency domain to remove drifts and to improve the signal to noise ratio. Single subject data were transformation into Talairach space [9], spatially smoothed with a minimal three dimensional 4 mm half-width Gaussian in order to reduce inter-subject anatomical variability, and then grouped using a hierarchical random effects analysis [8] and overall analysis of variance (ANOVA). The minimum significance level was set to $p < 0.05$ corrected for multiple comparison by using a cluster-size threshold adjustment, based on Forman et al. Monte Carlo stimulation approach (10), extended to 3D data sets using the threshold size plug-in Brain Voyager QX.

Percent signal change analysis: For the region of interest (ROI) signal magnitude analysis (Figure 1B), activation was sampled from the peaks of activation of the group ANOVA analysis (see Figure 1A): in the bilateral lateral-occipital complex (LOC), left pre-central sulcus (PCS), bilateral occipito-parietal sulcus (OccParS) and joining of the right occipito-parietal sulcus and anterior calcarine sulcus. Activation was then averaged across the four runs (using in each subject separately the peak voxel in a smoothed volume, after convolution with a Gaussian kernel of 4 mm full width at half maximum). The averaged percent signal change and standard errors were then calculated for each condition.

RESULTS

We directly compared a group of seven subjects learning to perform object recognition via soundscapes with a group of seven subjects learning arbitrary associations between sounds and object identity. This was done by ANOVA comparisons with a group factor (object via soundscapes; objects via associations) and conditions (vOICeObj; vOICeScr; AudObj). Figure 1A presents the direct contrast for areas showing significant differences between groups for the contrast of the vOICe objects as compared to the auditory controls that did not contain shape information (vOICeObj > vOICeScr, AudObj).

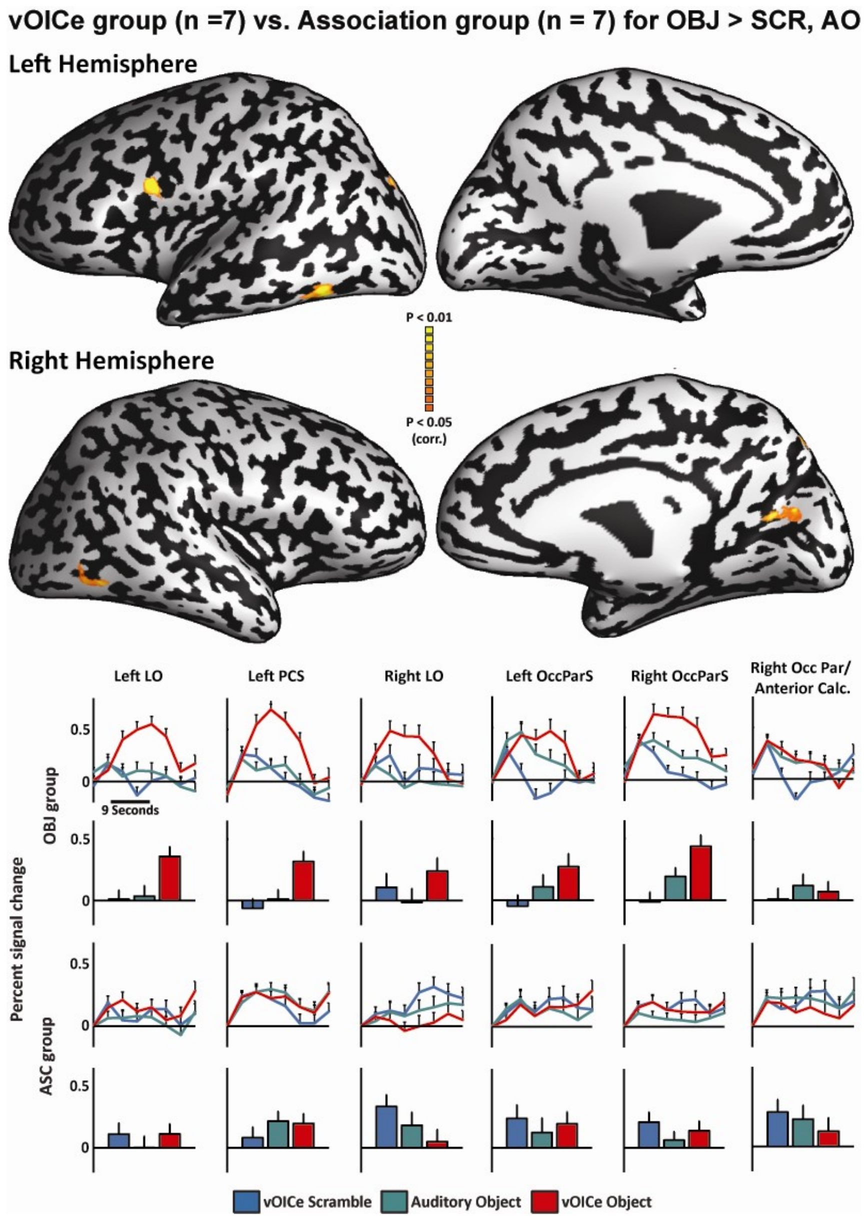

Figure 1. The neural network of sensory substitution object shape recognition. A. Activation for recognizing SSD objects (contrasted with SSD scrambled images and auditory objects; OBJ > SCR, AO) is compared by a RFX ANOVA analysis between the two groups, which differ in their ability to extract shape information from the SSD object stimuli. A group of seven subjects who learned to perform object recognition via shape extraction from soundscapes (OBJ group) is compared here to a group of seven subjects who learned the same soundscapes identities by arbitrary associations between sounds and object identity (ASC group). The most significant areas activated for this contrast (and thus the most critical structures which are specific to object shape recognition via soundscapes) were the left Pre-Central Sulcus (PCS) and the bilateral Lateral-Occipital cortex (LO). B. Time courses from each significant ANOVA cluster show that LO and PCS exhibit differential activation for vOICe objects versus the two other object types but only in the group who learned to extract relevant shape information.

The most significant areas for this comparison, and thus the most critical structures supporting object shape recognition via soundscapes, were the left Pre-Central Sulcus (PCS), and the bilateral Lateral-Occipital Complex (LOC). We also found activation in the occipito-parietal sulcus and the right posterior medial occipital cortex (Occipito-parietal – anterior calcarine junction, though this area exhibited a relatively weaker signal and less specificity, see Figure 1B). Time courses from each significant cluster (Figure 1B) indicated that LOtv in the ventral visual stream and PCS in the prefrontal cortex showed strong differential activation for vOICe objects versus the two other object types but only in the group who learned to extract relevant shape information. The occipito-parietal sulcus showed more selective activation for vOICe objects in the right hemisphere.

DISCUSSION

These results suggest that in addition to LOC, an entire cortical network is involved in the extraction of cross-modal shape information for object recognition, as opposed to identifying objects based on arbitrary associations of stimulus and identity. These regions include the left pre-central sulcus, the right occipito-parietal sulcus region and to some extent, the occipito-parietal sulcus/anterior calcarine region, in addition to the previously reported lateral occipital complex [4].

These results support the assumption that interactions between visual structures and a specific area in prefrontal cortex (PCS) might define the critical underlying mechanism for achieving object shape recognition via SSD soundscapes, similar to the mechanism evidenced for visual object recognition [11]. Areas in the lateral prefrontal cortex have been shown to match object identities across modalities [12, 13], and to have a cross-modal repetition suppression effect [14] for the combination of visual and haptic stimuli, suggesting a common neuronal basis for visuo-haptic integration of object identities (using fMRI adaptation design). Our findings suggest that at least part of this crossmodal platform is linked to a multisensory comparison of shape, and not to a more abstract association based solely on arbitrary association or its learning. This finding also supports the expansion of the metamodal shape network, which is engaged in deciphering shape regardless of input modality [4, 7], to a specific part of the prefrontal cortex, whereas at least another close-by area also encodes arbitrary associations between artificially determined (learned) visuo-auditory “objects” [13].

Interestingly, an additional area which is commonly found in cross-modal binding and learning with regard to object recognition, the intra-parietal sulcus [13, 14], does not appear, in this study, to be metamodally activated for processing object shape, and may thus be involved in binding multisensory object representations via familiarity and learning of the object identity.

These data contribute to the growing body of evidence that sensory substitution can cross-modally activate otherwise sensory 'specific' areas (see also [15-21] for studies supporting this notion with non-SSD approaches). Finally, this study also confirms that sensory substitution devices are powerful tools for studying key issues in brain research, such as multisensory interactions, brain plasticity, learning and object recognition.

ACKNOWLEDGMENTS

We thank Zohar Tal for help in commenting and editing the final draft of the paper. This work was supported in part by an R21-EY0116168 (to APL), the International Human Frontiers Science Program Organization Career Development Award, The Israel Science Foundation (grant number 1530/08); a European Union Marie Curie International Reintegration Grant (MIRG-CT-2007-205357); the Edmond and Lily Safra Center for Brain Sciences; and the Alon, Sieratzki, and Moscona funds (To AA). We would also like to thank the Hebrew University Hoffman Leadership and Responsibility Fellowship Program for its support (to ESA).

REFERENCES

[1] Meijer PB. An experimental system for auditory image representations. IEEE Trans Biomed Eng 1992; 39(2):112-21.

[2] Bach-y-Rita P, Kercel SW. Sensory substitution and the human-machine interface. *Trends Cogn. Sci.* 2003; 7(12):541-6.

[3] Renier L, Collignon O, Poirier C, Tranduy D, Vanlierde A, Bol A, Veraart C, De Volder AG. Cross-modal activation of visual cortex during depth perception using auditory substitution of vision. *Neuroimage* 2005; 26(2):573-80.

[4] Amedi A, Stern WM, Camprodon JA, Bermpohl F, Merabet L, Rotman S, Hemond C, Meijer P, Pascual-Leone A. Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat. Neurosci*. 2007; 10(6):687-9.

[5] Collignon O, Lassonde M, Lepore F, Bastien D, Veraart C. Functional Cerebral Reorganization for Auditory Spatial Processing and Auditory Substitution of Vision in Early Blind Subjects. *Cereb Cortex* 2006.

[6] Amedi A, Jacobson G, Hendler T, Malach R, Zohary E. Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cereb Cortex* 2002; 12(11):1202-12.

[7] Pascual-Leone A, Hamilton R. The metamodal organization of the brain. *Prog. Brain Res*. 2001; 134:427-45.

[8] Friston KJ, Holmes AP, Worsley KJ. How many subjects constitute a study? *Neuroimage* 1999; 10(1):1-5.

[9] Talairach J, Tournoux P. *Co-planar stereotaxic atlas of the human brain*. New York: Thieme; 1988.

[10] Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.* 1995; 33(5):636-47.

[11] Sehatpour P, Molholm S, Schwartz TH, Mahoney JR, Mehta AD, Javitt DC, Stanton PK, Foxe JJ. A human intracranial study of long-range oscillatory coherence across a frontal-occipital-hippocampal brain network during visual object processing. *Proc. Natl. Acad. Sci. USA* 2008; 105(11):4399-404.

[12] Meienbrock A, Naumer MJ, Doehrmann O, Singer W, Muckli L. Retinotopic effects during spatial audio-visual integration. *Neuropsychologia* 2007; 45 (3):531-9.

[13] Naumer MJ, Doehrmann O, Muller NG, Muckli L, Kaiser J, Hein G. Cortical Plasticity of Audio-Visual Object Representations. *Cereb Cortex* 2008.

[14] Tal N, Amedi A. Multisensory visual-tactile object-related network in humans: insights from a novel crossmodal adaptation approach. Experimental *Brain Research* 2009; 198(2-3):165-82.

[15] Beauchamp MS. See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr. Opin. Neurobiol.* 2005; 15(2):145-53.

[16] Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ. Functional imaging of human crossmodal identification and object recognition. *Exp. Brain Res*. 2005; 166(3-4):559-71.

[17] James TW, Humphrey GK, Gati JS, Servos P, Menon RS, Goodale MA. Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia* 2002; 40(10):1706-14.

[18] Sathian K. Visual cortical activity during tactile perception in the sighted and the visually deprived. *Dev. Psychobiol.* 2005; 46(3):279-86.

[19] Pascual-Leone A, Amedi A, Fregni F, Merabet LB. The plastic human brain cortex. *Annu. Rev. Neurosci.* [Review] 2005; 28:377-401.

[20] Zangaladze A, Epstein CM, Grafton ST, Sathian K. Involvement of visual cortex in tactile discrimination of orientation. *Nature* 1999; 401(6753):587-90.

[21] Hein G, Doehrmann O, Muller NG, Kaiser J, Muckli L, Naumer MJ. Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J. Neurosci.* 2007; 27(30):7881-7.

Received: February 6 2011
Revised: February 17 2011
Accepted: February 18 2011.