Unraveling multisensory integration: patchy organization within human STS multisensory cortex

Michael S Beauchamp[1], Brenna D Argall[1], Jerzy Bodurka[2], Jeff H Duyn[3] & Alex Martin[1]

Although early sensory cortex is organized along dimensions encoded by receptor organs, little is known about the organization of higher areas in which different modalities are integrated. We investigated multisensory integration in human superior temporal sulcus using recent advances in parallel imaging to perform functional magnetic resonance imaging (fMRI) at very high resolution. These studies suggest a functional architecture in which information from different modalities is brought into close proximity via a patchy distribution of inputs, followed by integration in the intervening cortex.

The human superior temporal sulcus multisensory area (STS-MS) is important for integrating auditory and visual information about objects, speech, letters and other behaviorally relevant stimuli[1–4]. Electrophysiological recording studies from macaque monkeys demonstrate that individual neurons in monkey STS may respond only to auditory stimuli, only to visual stimuli, or both to auditory and to visual stimuli[5,6]. Although it is reasonable to assume that similar neuronal response properties exist in human STS-MS, there has been no direct evidence for this. Additionally, electrophysiological and functional neuroimaging studies to date have provided no information on the topographic organization of these different types of neurons.

One possibility is that the STS-MS is organized as a homogeneous mixture of auditory, visual and auditory-visual neurons. Arguing against this idea is the observation from tracer injection studies that auditory and visual projections to monkey STS lie in non-overlapping domains[7]. This patchy organization is on a scale of 1–2 mm (ref. 8). Owing to technical limitations, standard-resolution fMRI uses voxels that are too large (40–70 mm^3) to observe fine structure within cortical areas. Recent advances in multichannel MRI receivers[9] and whole-brain surface coil phased arrays[10] provide improved signal-to-noise ratio and permit the acquisition of high-resolution fMRI data with significantly more flexibility than single surface coils[11,12], making them ideally suited to study the STS-MS.

We mapped the STS-MS in human subjects using standard-resolution fMRI and either videos of tools (for example, a hammer making a hammering motion), recordings of

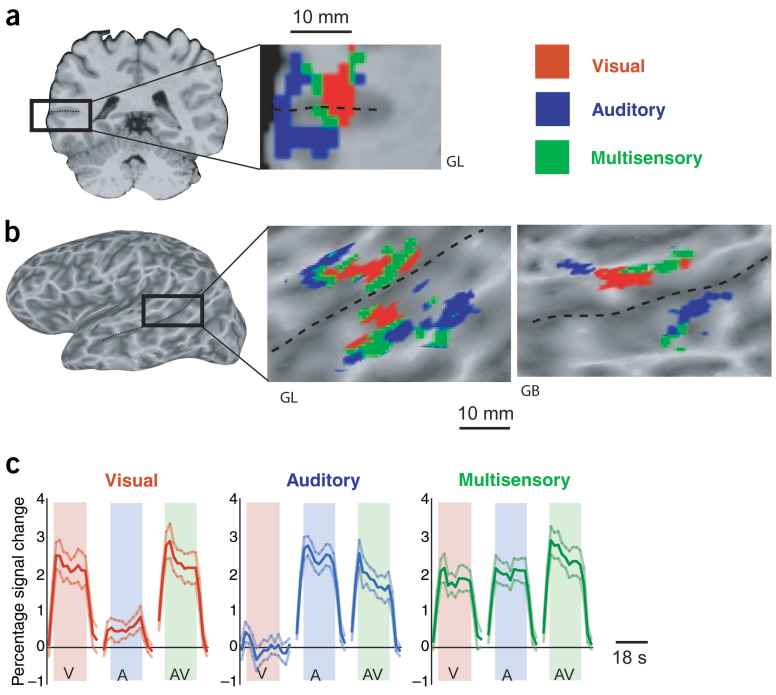

Figure 1 Patchy organization within the STS-MS. (**a**) Coronal section with enlargement of the left STS (dashed line). Colors show relative response to unisensory visual (V) and auditory (A) tools. Orange (visual patches): V > A, $P < 0.05$. Blue (auditory patches): A > V, $P < 0.05$. Green (multisensory patches): A = V, $P < 0.05$. Two-letter code (GL) indicates subject identity. (**b**) Lateral view of the left hemisphere of an inflated cortical surface model, with enlargement showing the STS-MS in two subjects. Same color scale as in **a**. (**c**) Average MR time series across subjects ($n = 8$). Three graphs showing the response in visual (left), auditory (middle) and multisensory (right) patches to the three stimulus types (pink shaded region, V, response to visual tools; blue shaded region, A, response to auditory tools; green shaded region, AV, response to multisensory tools) and fixation baseline (non-shaded regions). Thick line, mean response; thin line, s.e.m.

[1]Laboratory of Brain and Cognition and [2]Functional MRI Facility, National Institute of Mental Health, and [3] Section on Advanced MRI, Laboratory of Functional and Molecular Imaging, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA. Correspondence should be addressed to M.S.B. (mbeauchamp@nih.gov).

Published online 10 October 2004; doi:10.1038/nn1333

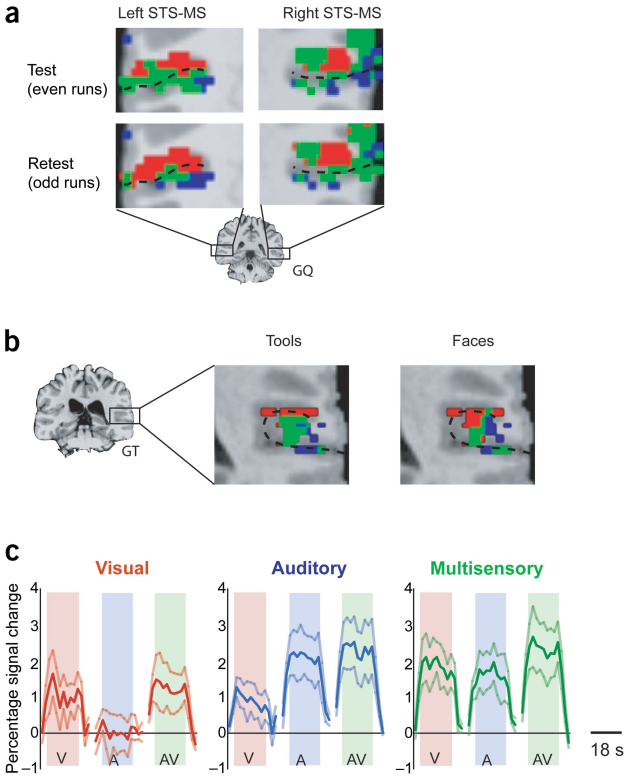

Figure 2 Reliability of patchy organization within the STS-MS. (**a**) Test-retest reliability in a single subject with the tool stimulus set (same color scale as in **Fig. 1**). (**b**) Test-retest reliability with different stimulus sets (tools and faces/voices). (**c**) The response to face/voice stimuli in patches defined using tool stimuli, averaged across subjects ($n = 5$). Each graph shows the response in a single patch type to stimuli indicated by shaded bars (V, videos of faces; A, recordings of voices; AV, faces + voices) relative to fixation baseline. Thick line, mean response; thin line, s.e.m.

tools (for example, "bang-bang-bang") or simultaneous videos and recordings of tools (see **Supplementary Videos 1–6** and refs. 1,13). In a second experimental session, the same subjects were scanned at high resolution (1.6 mm × 1.6 mm× 1.6 mm) using a phased-array coil to investigate the fine structure of the STS-MS (see **Supplementary Methods**). Across subjects, STS-MS was slightly larger in right than left hemispheres (1,466 mm^3 versus 1,224 mm^3, not significant.) As shown in volume (**Fig. 1a**) and surface representations (**Fig. 1b**), the STS-MS contained contiguous patches of voxels with different response properties. Among STS-MS voxels, 44% were located in auditory patches that responded more to unimodal auditory (A) than unimodal visual (V) stimuli; 30% were located in visual patches that preferred unimodal visual stimuli to unimodal auditory stimuli; and 26% were found in multisensory patches that responded equally to unimodal auditory and visual stimuli. This patchy organization was observed on both the upper and lower banks of the STS.

A number of statistical artifacts can give rise to organized structure in cortical maps. Thus, we carried out additional analyses and experiments to rule out artifacts and to gain additional insight into the organization of the STS-MS.

First, we examined the average MR response to A, V and auditory-visual (AV) stimulus conditions in the three types of cortical patches (**Fig. 1c**). Because only the unisensory (A and V) responses were used to define the patches, the AV response represents a critical test of the validity of the results. Instead of the random variation that would be expected if the patchy organization were an artifact of the analysis, systematic differences in the AV response were observed. In visual patches, the AV response amplitude equaled the V response, whereas in auditory patches the AV response equaled the A response, suggesting little sensitivity to the non-preferred modality. In contrast, multisensory patches responded more to AV than to either A or V stimuli, highlighting the multisensory integration occurring in these patches (mean response to AV condition, 2.5%; mean response to A and V conditions, 1.9%; $P = 0.008$). Although the response was greatest in the AV condition, the response was less than the summed A and V responses (2.5% versus 3.9%, $P = 0.003$).

Second, we measured the test-retest reliability of the patchy organization. For one individual subject (**Fig 2a**), the spatial correlation between test and retest maps was $r = 0.92$ across 250 voxels ($P < 10^{-40}$). Consistent test-retest maps were observed in all subjects (mean $r = 0.82$, minimum $r = 0.58$, P for least reliable subject $< 10^{-6}$).

If the patchy fMRI activation reflects an underlying anatomical organization, it should be consistent across stimulus sets. Additional stimuli were created of videos of unfamiliar human faces (V), recordings of voices (A) and simultaneous faces and voices (AV). The activation map constructed from the face stimuli showed a patchy distribution that appeared qualitatively similar to that observed with tool stimuli (**Fig. 2b**). The response to face stimuli was measured in patches independently defined by the tool stimuli (**Fig. 2c**). Voices evoked a stronger response than faces in auditory patches ($P = 0.02$), faces evoked a stronger response than voices in visual patches ($P = 0.0003$), and multisensory (face + voice) stimuli evoked a stronger response than voices or faces in multisensory patches (2.9% versus 2.3%, $P = 0.04$). To better understand the influence of stimulus category on the patchy organization, a two-factor analysis of variance (ANOVA) was performed on the MR response within each patch type. The first factor was the stimulus modality (A, V or AV) and the second factor was stimulus category (tools or faces). The replications consisted of determining the percentage MR signal change in each of the five subjects who were scanned using both stimulus sets. Unsurprisingly, stimulus modality had a significant effect in visual patches ($F_{2,29} = 5.1$, $P = 0.01$) and auditory patches ($F_{2,29} = 7.6$, $P = 0.003$). Stimulus category had a significant response in visual patches ($F_{1,29} = 6.5$, $P = 0.02$). The response was significantly greater to tools than faces (likely because of the greater amount of visual motion contained in the tool videos). However, there was no significant interaction between modality and category in any of the patch types (interaction in visual patches, $F_{2,29} = 0.12$, $P = 0.9$; auditory patches, $F_{2,29} = 0.39$; $P = 0.7$; multisensory patches, $F_{2,29} = 0.11$, $P = 0.9$). Therefore, the modality preference of the different patch types was independent of stimulus category, as would be expected if the patches reflect anatomically defined compartments receiving segregated sensory input (see **Supplementary Discussion**). An additional two-factor repeated-measures ANOVA was performed on the face data, with the first factor being the stimulus modality (A, AV or AV) and the second factor the patch type (as defined using the tool stimuli). There was a weak but significant interaction between stimulus modality and patch type ($F_{4,44} = 2.7$, $P = 0.048$), indicating that the three patch types differed in their responses to the three types of face stimuli. We conclude that the underlying patchy organization in the STS-MS was preserved across stimulus sets.

As assessed using standard-resolution fMRI, the STS-MS appeared to consist of a homogenous region of cortex that responded similarly to auditory and visual stimuli. At tenfold higher resolution, a patchy organization within the STS-MS was revealed, with different patches responding maximally to auditory, visual or auditory-visual stimuli. Because of the reliability of the observed patchy organization, statistical artifact can be ruled out as the cause. Although fMRI does not directly measure neuronal activity, simultaneous electrophysiological and fMRI studies have shown high correlation between the local field potential and the response seen on blood oxygen level–dependent (BOLD) fMRI[14]. BOLD fMRI is especially sensitive to the synaptic inputs into a given cortical region. Therefore, the separate auditory and visual patches that were observed are likely to reflect concentrations of individual neurons that are receiving primarily auditory or visual inputs. The intervening multisensory patches that show an enhanced response to auditory-visual stimuli (although not a superadditive response; see **Supplementary Discussion**) are likely to reflect concentrations of multisensory auditory-visual neurons.

A model for cortical multisensory organization suggested by our data is that auditory and visual inputs arrive in the STS-MS in separate patches, followed by integration in the intervening cortex. Combined with physiological data from other methodologies (**Supplementary Discussion**; ref. 15), the functional organization described here provides important clues about the cortical architecture underlying auditory-visual integration. Moreover, it may reflect a general strategy used to integrate information arising in different processing streams throughout neocortex.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We thank C. Senior for providing the face stimuli and R.W. Cox and Z.S. Saad for their continued development of AFNI and SUMA.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 7 July; accepted 23 September 2004
Published online at http://www.nature.com/natureneuroscience/

1. Beauchamp, M.S., Lee, K.E., Argall, B.D. & Martin, A. *Neuron* **41**, 809–823 (2004).
2. Calvert, G.A. *Cereb. Cortex* **11**, 1110–1123 (2001).
3. Van Atteveldt, N., Formisano, E., Goebel, R. & Blomert, L. *Neuron* **43**, 271–282 (2004).
4. Wright, T.M., Pelphrey, K.A., Allison, T., McKeown, M.J. & McCarthy, G. *Cereb. Cortex* **13**, 1034–1043 (2003).
5. Hikosaka, K., Iwai, E., Saito, H. & Tanaka, K. *J. Neurophysiol.* **60**, 1615–1637 (1988).
6. Benevento, L.A., Fallon, J., Davis, B.J. & Rezak, M. *Exp. Neurol.* **57**, 849–872 (1977).
7. Seltzer, B. & Pandya, D.N. *J. Comp. Neurol.* **343**, 445–463 (1994).
8. Seltzer, B. *et al. J. Comp. Neurol.* **370**, 173–190 (1996).
9. Bodurka, J. *et al. Magn. Reson. Med.* **51**, 165–171 (2004).
10. De Zwart, J.A. *et al. Magn. Reson. Med.* **51**, 22–26 (2004).
11. Cheng, K., Waggoner, R.A. & Tanaka, K. *Neuron* **32**, 359–374 (2001).
12. Goodyear, B.G. & Menon, R.S. *Hum. Brain Mapp.* **14**, 210–217 (2001).
13. Beauchamp, M.S., Lee, K.E., Haxby, J.V. & Martin, A. *Neuron* **34**, 149–159 (2002).
14. Logothetis, N.K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. *Nature* **412**, 150–157 (2001).
15. Wallace, M.T., Ramachandran, R. & Stein, B.E. *Proc. Natl. Acad. Sci. USA* **101**, 2167–2172 (2004).