Published in final edited form as:

Behav Brain Res. 2013 February 1; 238: 265–272. doi:10.1016/j.bbr.2012.10.011.

Early Sensory Cortex is Activated in the Absence of Explicit Input During Crossmodal Item Retrieval: Evidence from MEG

Ajay S. Pillai*, Jessica R. Gilbert*, and Barry Horwitz

Brain Imaging and Modeling Section, National Institute on Deafness and other Communication Disorders, National Institutes of Health, Bethesda MD 20892.

Abstract

Crossmodal associations form a fundamental aspect of our daily lives. In this study we investigated the neural correlates of crossmodal association in early sensory cortices using magnetoencephalography (MEG). We used a paired associate recognition paradigm in which subjects were tested after multiple training sessions over a span of four weeks. Subjects had to learn 12 abstract, nonlinguistic, pairs of auditory and visual objects that consisted of crossmodal (visual-auditory, VA; auditory-visual, AV) and unimodal (visual-visual, VV; auditory-auditory, AA) paired items. Visual objects included abstract, non-nameable, fractal-like images, and auditory objects included abstract tone sequences. During scanning, subjects were shown the first item of a pair (S1), followed by a delay, then the simultaneous presentation of a visual and auditory stimulus (S2). Subjects were instructed to indicate whether either of the S2 stimuli contained the correct paired associate of S1. Synthetic Aperture Magnetometry (SAMspm), a minimum variance beamformer, was then used to assess source power differences between the crossmodal conditions and their corresponding unimodal conditions (i.e., AV-AA and VA-VV) in the beta (15–30 Hz) and low gamma frequencies (31–54 Hz) during the S1 period. We found greater power during S1 in the corresponding modality-specific association areas for crossmodal compared with unimodal stimuli. Thus, even in the absence of explicit sensory input, the retrieval of well-learned, crossmodal pairs activate sensory areas associated with the corresponding modality. These findings support theories which posit that modality-specific regions of cortex are involved in the storage and retrieval of sensory-specific items from long-term memory.

Keywords

crossmodal association; long-term memory; magnetoencephalography

INTRODUCTION

The ability to form associations and subsequently use them is an integral aspect of long-term memory processes. These associations can be made automatically or explicitly learned, both

© 2012 Elsevier B.V. All rights reserved.

Corresponding Author: Dr. Jessica R. Gilbert Address: Brain Imaging and Modeling Section National Institute on Deafness and other Communication Disorders (NIDCD) National Institutes of Health (NIH) 10 Center Drive, room 5D39 Bethesda, Maryland 20892

Phone: 301-594-7757 Fax: 301-480-5625 jessica.gilbert@nih.gov.

*ASP and JRG contributed equally to this work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of interest: The authors declare that they have no conflict of interest.

within and across sensory modalities. While studies of long-term memory have typically focused on within-modality (i.e., unisensory) learning [1,2], more recent studies have taken an interest in crossmodal associative learning [3-5]. Crossmodal associations are of particular interest because they have been shown to confer behavioral advantages compared with unisensory identification [6]. Auditory objects can evoke mental imagery, as in the case of the spoken word 'church' evoking the picture of a church in one's mind. This can also happen with the visual presentation of the written word 'church'. While these are quite disparate sensory inputs, either can lead to fast and efficient object recognition. While several recent studies have focused on the learning of crossmodal associations (e.g., [3,4]), few studies have examined the retrieval of well-learned (i.e., long-term), abstract paired associates from unimodal (i.e., auditory-auditory, visual-visual) compared with crossmodal (i.e., auditory-visual, visual-auditory) paired items.

Several theories exist which account for complex memory formation and subsequent recall. Traditionally, memory processes were considered to be separate from perceptual processes [7]. However, more recent evidence suggests there is a symbiotic relationship between perceptual processes and processes involved in both encoding and retrieval of long-term memories [8-11]. For example, Damasio [12] suggested that modality-specific regions of cortex store perceptual and motor experiences and that reactivation of these regions by higher-order areas is crucial for successful memory retrieval. Elaborating on this idea, Fuster [10] suggested that a memory code, which is formed by experience, is fundamentally a relational coderepresented by sparse, distributed neuronal networks which can overlap. The degree and nature of inter-regional connectivity is determined by the type of sensory experience (e.g., visual, tactile, auditory) and the type of memory. More complex memories then are formed by connecting smaller networks. Such hierarchical organization can give rise to a vast variety of memory networks due to various combinatorial possibilities.

Building on these ideas, Damasio has proposed the existence of convergence-divergence zones (CDZs) [11,12], which can be thought of as neural networks that store associations about how characteristics encoded within sensory and motor (i.e., modality-specific) cortices must be brought together to form complete percepts of memories. In the case of crossmodal associations, CDZs should activate the corresponding associated modality even in the absence of sensory input. Such examples have been found in a variety of domains. For example, silent lipreading has been shown to activate auditory areas [13]. Similarly, it has been shown that reading words strongly associated with smell can activate olfactory regions [14] and those strongly associated with sounds can evoke auditory responses [15]. Previous animal studies have also shown the existence of mnemonic neural responses to visual long-term paired associates [1,16], visual-auditory paired associates [17], and visual-tactile associates [18]. In the same way, recent evidence using fMRI in humans suggests learning to extract object-shape information from learned auditory cues can activate regions responsive to tactile-visual information [19], and learning to manipulate novel objects produces activation in motor cortices during visual object identification [20]. These findings suggest that long-term memory retrieval engages areas originally involved in the sensory processing of the stimulus. Thus, we would predict that presenting a visual object which has been paired with an auditory object should produce early activation in auditory association areas, even in the absence of direct sensory input. In the same way, presentation of an auditory object that has been paired with a visual input should produce early activation in visual association areas.

A few previous studies have addressed crossmodal versus unimodal paired associate learning-related activity. For example, Tanabe et al. [3] explored brain activity during the learning of abstract crossmodal and unimodal paired stimuli using fMRI. They found activation within the associated stimulus modality (i.e., visual cortex activity for auditory

paired associates and auditory cortex activity for visual paired associates) during the delay period between the presentation of the first and second stimulus as learning proceeded. However, this study included a long delay period (16 s) between the first and second stimulus presentation in order to resolve the slow BOLD signal response. Even with the long delay, it is difficult to disentangle stimulus-specific responses to the first stimulus, processes involved in working memory maintenance, and those processes involved in the retrieval of the paired associate. A second fMRI study by Naumer et al. [4] also showed recruitment of auditory and visual processing areas (i.e., superior temporal sulcus and fusiform gyrus) for artificial AV paired associates following a day of training. Finally, Butler & James [5] showed within-modality 'reactivation' within primary auditory and visual cortices using fMRI, but not crossmodal activation in the absence of sensory input. That is, they showed enhanced activity in visual cortex for visual stimuli from unimodal compared with crossmodal pairs (i.e., visual-visual compared to visual-auditory), and enhanced activity in auditory cortex for auditory stimuli from unimodal compared with crossmodal pairs (i.e., auditory-auditory compared to auditory-visual). However, the use of fMRI makes it hard to disentangle fine-grained temporal effects as the resolution of fMRI is on the order of seconds. Thus, it is still unclear whether well-learned abstract auditory and visual pairs activate the corresponding crossmodal sensory cortex early, in the absence of explicit sensory input (i.e., enhanced activity in auditory cortex during a visual stimulus that has been previously paired with an auditory stimulus, or enhanced activity in visual cortex during an auditory stimulus that has previously been paired with a visual stimulus), though a previous MEG study does suggest early network divergence (by ~500 ms after stimulus presentation) for crossmodal compared with unimodal paired items [21].

In the current study, we therefore utilized the temporal benefits of MEG to explore differences in neural activity during crossmodal (i.e., AV and VA) compared with unimodal (i.e., AA and VV) paired associate recall. We employed a protocol in which abstract, nonlinguistic visual and auditory stimuli were paired over several learning sessions prior to scanning. We then assessed source-power differences between well-learned item pairs having either a crossmodal associate (i.e., auditory-visual or visual-auditory) or a unimodal (i.e., auditory-auditory or visual-visual) associate. Specifically, we predicted that crossmodal stimuli should activate the associated stimulus modality in the absence of explicit sensory input. MEG is an ideal tool to address this given its exquisite temporal resolution (i.e., on the order of ms). Therefore, given our paradigm and imaging modality, we directly assessed whether the associated sensory cortex is activated for crossmodal paired associates compared to within-modality unimodal associates, prior to direct sensory inputs (i.e., during the presentation of the first stimulus within a pair: S1). This allowed us to control for stimulus specific processing (e.g., both VA and VV have matched visual inputs during S1), while determining whether there was increased activity in the associated stimulus modality for crossmodal compared with unimodal paired items.

MATERIALS AND METHODS

Participants

This study was approved by the National Institutes of Health ethics committee and performed in accordance with ethical standards laid down in the 1964 Declaration of Helsinki. Data were collected from 18 subjects (10 females; Mean Age=27.5 years, Range=23-40). All subjects had self-reported normal hearing and normal or corrected-to-normal vision. Signed informed consents were obtained from all subjects prior to study participation, and all subjects were compensated for their time. For the MEG data presented in this manuscript, we set a performance criterion of 75% correct across all trial types, as we were interested in crossmodal responses to well-learned pairs of stimuli. This resulted in 13

participants who met our criteria and whose data were used for subsequent analyses (see supplementary material for performance data from all participants).

Experimental Design

We employed a paired associate protocol in which 12 nonlinguistic, abstract, visual and auditory stimuli were used (see supplementary material for example stimuli). These included four types of paired items that subjects were required to commit to memory containing either crossmodal (auditory-visual, AV, N=3; visual-auditory, VA, N=3) or unimodal (auditory-auditory, AA, N=3; visual-visual, VV, N=3) paired items (see Figure 1A).

Subjects were trained in 3 separate sessions lasting approximately 1.5 hours each. Each of the training sessions were separated by a week (days 1, 8, and 15). One week following the last training session (day 22), subjects were scanned using MEG while they made correct/incorrect judgments on pairs of items. The MEG scanning session lasted approximately 20 minutes. During scanning, subjects were first presented with stimulus one (S1) for 1 s (which was either a visual or auditory stimulus depending on the trial type), and after a delay of 1 s they were presented with stimulus two (S2) for 1 s, which was the simultaneous presentation of a visual and auditory stimulus. A total of 168 trials were presented during scanning, with 42 trials of each stimulus type. Subjects were instructed to indicate via button press as quickly as possible whether either of the modalities of the compound S2 contained the correct paired associate of the given S1. Correct pairings were presented on 50% of trials. Trial type and congruency were randomized across the entire run.

Visual stimuli included abstract non-linguistic, fractal-like images. All S1 visual objects were blue-green fractal-like images, and S2 visual objects were green-yellow fractal-like images. Similarly, the auditory objects were complex abstract tone sequences 1 second in length. Each 1 second stimulus was composed of 4 tones of 250 ms duration each, which were separated by a base frequency corresponding to a pitch difference of 50 mels. All the tones used to create S1 sequences were between 350 and 800 mels (indicated by the blue waveform in Figure 1). Similarly all tones used to create S2 sequences were between 950 and 1400 mels (indicated by the red waveform in Figure 1). This separation of at least 150 mels between S1 and S2 introduced a clear perceptual separation in all subjects (confirmed verbally during the training session), with all the S1s sounding lower in pitch and the S2s sounding higher in pitch. In addition, the base frequency (i.e., pitch) of the first segment (250 ms) of each of the tone sequences used was unique. So, in principle, the subjects had enough information in the first 250 ms to differentiate between any two tone sequences.

Data Acquisition

MEG data were recorded at a 600 Hz sampling rate with a bandwidth of 0-150 Hz using a CTF 275 MEG system (VSM MedTech Ltd., Canada) composed of a whole-head array of 275 radial 1st order gradiometer channels housed in a magnetically-shielded room (Vacuumschmelze, Germany). Synthetic 3rd gradient balancing was used to remove background noise on-line. Fiducial coils were placed on the nasion, left preauricular, and right preauricular sites of each participant, and were energized prior to stimulus presentation to localize the participant's head with respect to the MEG sensors. Total head displacement was measured and could not exceed 5 mm for inclusion in the source analyses. Individual fiducial locations were then used to coregister each participant's data to high-resolution T1-weighted anatomical images acquired in a separate scanning session with a 3-Tesla whole-body scanner (GE Signa, Milwaukee, WI).

MEG Source Analysis

A beamformer analysis was used to localize sources of activity during the testing session. Beamformers are spatial filters that estimate the underlying spatial locations of MEG activity recorded by the sensors. A new implementation of synthetic aperture magnetometry (SAMspm), a minimum variance beamformer which computes actual statistics across trials [22], was used to assess source power differences between the corresponding crossmodal (AV, VA) and unimodal (AA, VV) conditions. SAMspm is an adaptive beamforming technique which calculates power across the brain on a voxel by voxel basis within a specified frequency band and time window. We calculated the power ratio for each subject from S1 onset for 1 s (i.e., entire S1 period), directly contrasting crossmodal (i.e., 'active') to unimodal (i.e., 'control') stimuli (i.e., VA-VV and AV-AA) using a 5 mm voxel resolution, in the following frequency bands: beta (15-30Hz), low gamma (31-54Hz), and high gamma (55-90 Hz). Beta and gamma band synchrony (i.e., power increase) has been implicated in a wide variety of cognitive and perceptual phenomena, thus we focused on these frequencies in our analysis (e.g., [23-25]). In addition, since we were interested in source-level differences during S1 for crossmodal compared with unimodal paired associates at a time when sensory inputs were matched (thus controlling for unimodal sensory inputs), we did not have adequate samples in the lower frequencies (i.e., theta and alpha) to construct stable covariance matrices. For example, Van Veen et al. [26] suggest that the minimum number of effective samples must be greater than three times the number of sensors. Since the covariance matrix is a weighted estimate of activity across the MEG sensors (here, 275), more data will yield a more reliable source estimate. Thus, there are not enough effective samples in the 1 s window of interest to localize theta and alpha frequencies. Therefore, we did not assess source-power differences in frequencies below beta. For each voxel, SAMspm applies the beamformer weights to the unaveraged signal, thus giving an estimate of the source time-series. The time series is then parsed into 'active' and 'control' time segments, and the band-limited source power is accumulated over each segment. The variance in source power between the active and control segments is then used to produce an F ratio between the active and control conditions. We utilized only trials where subjects responded correctly as to whether the compound S2 stimulus contained the correct paired associate of the presented S1 stimulus. The power ratios calculated at each voxel then give the relative change in power at a given voxel between the two conditions. This contrast controls for stimulus-specific processing, as in both cases the sensory input is matched. The only difference is that for the 'active' period the associated stimulus is crossmodal, while for the 'control' period the associated stimulus is unimodal. A positive value then indicates relatively greater power in the crossmodal condition (i.e., AV or VA respectively), and a negative value indicates relatively greater power in the unimodal condition (i.e., AA or VV respectively) for a given voxel. Each participant's data were then normalized and converted to Talairach space using AFNI [27] for group-level comparisons. To assess statistical significance for each contrast, we utilized a 1-sample t-test to compare the group-level results from our comparison conditions to zero, thresholded at $p < 0.01$ corrected (we used a non-parametric, random permutation test to correct for multiple comparisons).

In addition, because we were interested in the relative timing of differences between crossmodal and unimodal trials in early sensory cortices that showed a significant effect, we used a sliding window technique to assess power differences in shorter, overlapping time windows. These sliding window analyses allowed us to determine whether power differences (here, defined as greater power for crossmodal relative to unimodal pairs) in the crossmodal sensory cortices occurred early or late during our 1 s window of interest. Using the original covariance matrix and weights files constructed from S1 onset to 1 s post-onset for both the AV-AA and VA-VV comparisons, we used a 200 ms window beginning 100 ms

prior to S1 onset, slid every 100 ms, directly contrasting AV-AA and VA-VV pairs in both the beta and low gamma frequencies (there were no consistent effects in high gamma). This resulted in 11 sampling windows from -100 to 1100 ms relative to S1 onset. Using masks of the regions in early visual and auditory cortices identified in our beta and low gamma frequency comparisons of AV-AA and VA-VV as regions of interest, we then assessed the mean power difference in each region across our 11 time samples. One-sample t-tests were used to determine whether each time sample showed a significant ($p < 0.05$) difference compared to zero.

RESULTS

Behavioral Results

Thirteen participants met our behavioral criteria of 75% correct across all trial types. For these participants, we found a significant effect of paired associate type on accuracy [$F(3,12)=17.7$, $p < 0.01$] and reaction times [$F(3,12)=20.1$, $p < 0.01$]. Using a Bonferroni correction, paired-sample t-tests showed that participants were significantly ($p < 0.05$ corrected) more accurate for VV pairs (Mean=39.5/42=94.1%) than for all other pairs of items (VA Mean=37.2/42=88.6%; AV Mean=34.6/42=82.4%; AA Mean=32.2/42=76.7%) and for VA pairs than AA pairs (see Figure 2). In addition, participants were significantly faster in responding to VV pairs (Mean=1208.6 ms) than all other pairs of items (VA Mean=1497.9 ms; AV Mean=1434.7 ms; AA Mean=1572.8 ms). Here, reaction time differences should have little bearing on the source analysis, as we assessed differences between crossmodal and unimodal stimuli during S1 only.

Source Localization

We localized source power differences between crossmodal (i.e., AV and VA respectively) and unimodal (i.e., AA and VV respectively) stimuli in the beta and gamma frequencies (low and high gamma separately) using SAMspm. We hypothesized that crossmodal stimuli should produce activation (i.e., greater power) in sensory-specific regions of cortex of the associated stimulus modality, even in the absence of sensory input. Thus, we predicted that AV-AA should produce greater activation in visual cortex, and that VA-VV should produce greater activation in auditory cortex. In both beta and low gamma frequencies we found support for this hypothesis. Source-level findings in high gamma did not show consistent effects in sensory-specific cortices for both the AV-AA and VA-VV comparisons. Thus, we have not presented the results here, but have made them available as supplementary material. Results for the other frequency bands are presented separately below.

Beamformer: Beta Band (15-30 Hz)—In the beta frequency, we found several regions showing differences between crossmodal AV and unimodal AA paired stimuli during the S1 period (see Table 1). Within the visual ventral stream, we found two regions showing greater power for AV compared with AA stimuli during the S1 period. The first region was located in the Cuneus bilaterally (see Figure 3A, left). The second region was located in the right fusiform gyrus (see Figure 3A, right). Our sliding window analysis indicated that these differences occurred early during the S1 period (see Figure 4). Within Cuneus, significant ($p < 0.05$) effects occurred within a window from 200-400 ms post-S1 onset; within the fusiform gyrus, significant effects occurred in a window from 100-300 ms post-S1 onset.

In the beta frequency we also found several regions showing differences between crossmodal VA and unimodal VV paired stimuli during the S1 period (see Table 1). Within auditory association areas, we found a single region showing greater power for VA compared with VV stimuli located in the right superior temporal gyrus (see Figure 3B). In our sliding window analysis, no time sample showed a significant ($p < 0.05$) effect, however

a trend toward significance ($p=0.07$) was identified in a window from 0-200 ms post S1-onset, and again at 300-500 ms and 500-700 ms post S1-onset (see Figure 4).

Beamformer: Low Gamma Band (31-54 Hz)—In the low gamma frequency, we found several regions showing differences between crossmodal AV and unimodal AA paired stimuli during the S1 period (see Table 2). Within the visual ventral stream, we found a single region showing greater power for AV compared with AA stimuli located in the right middle occipital gyrus (see Figure 5A). Our sliding window analysis indicated that significant ($p<0.05$) effects occurred within the first 100 ms post-S1 onset (continuing to 200 ms post-onset), and again in a window from 300-500 ms post-S1 onset (see Figure 6).

In the low gamma frequency we also found several regions showing differences between crossmodal VA and unimodal VV paired stimuli during the S1 period (see Table 2). Within auditory association areas, we found a single region showing greater power for VA compared with VV stimuli located in the right posterior superior temporal gyrus (see Figure 5B). Our sliding window analysis indicated that significant ($p<0.05$) effects occurred within the first 100 ms post-S1 onset (continuing to 400 ms post-onset), and again in a window beginning at 700-900 ms post-S1 onset (continuing to 1100 ms post-onset) (see Figure 6).

DISCUSSION

We utilized the temporal benefits of MEG to determine whether well-learned crossmodal paired associates produce activation within the associated sensory modality, even in the absence of explicit sensory input. That is, whether the presentation of an auditory stimulus could produce early activation in visual cortex, if that auditory stimulus had been previously paired with a visual stimulus. Likewise, could presentation of a visual stimulus produce early activation in auditory cortex, if it had been previously paired with an auditory stimulus? To answer this, we asked subjects to learn 12 abstract, non-linguistic pairs of items that included crossmodal (i.e., AV and VA) and unimodal (i.e., AA and VV) paired stimuli over 3 sessions each a week apart. A week following the last training session (i.e., at day 21), subjects were scanned using MEG while they performed a correct/incorrect judgment on pairs of items. MEG was used so that we could clearly differentiate whether early differences occurred in the associated crossmodal modality. We looked at source-power differences during the S1 period when the associated stimulus was crossmodal (i.e., AV or VA) compared to unimodal (i.e., AA or VV). Our results demonstrate greater power in the corresponding sensory modality for crossmodal compared with unimodal stimuli in both the beta and low gamma frequencies. That is, when an auditory stimulus was paired with a crossmodal visual stimulus, we found greater power in both beta and low gamma frequencies in visual association areas. Conversely, when a visual stimulus was paired with a crossmodal auditory stimulus, we found greater power in both beta and low gamma frequencies in auditory associations areas.

Previous studies have suggested that well-learned, linguistic crossmodal associations produce activation in the same sensory modality in the absence of explicit input (e.g., [13, 14]). Here we extend these findings to show that pairing abstract, non-linguistic visual and auditory stimuli leads to increased power in the associated stimulus modality, even in the absence of explicit input. That is, well-learned crossmodal paired stimuli appear to activate the corresponding sensory modality, even in the absence of linguistic associations. Given that animal studies have shown the existence of mnemonic neural responses to visual-auditory paired associates [17], these basic mechanisms have important implications for visual-auditory paired associations from a range of stimulus classes, perhaps even including the support of human language. We also show that this difference arises very early, during the presentation of the first stimulus of a pair (i.e., 1 s period). Given that our task involved

correct/incorrect judgments of pairs of items presented sequentially, we suggest that subjects are retrieving the paired associate of S1 from long-term memory during this period. Our finding thus indicate that long-term memory storage engages areas originally involved in the sensory processing of the stimulus.

The finding that retrieval of a paired associate from long-term memory activates sensory-specific cortices supports theories that advocate for what has been termed ‘grounded cognition’ (e.g., [8-11]). That is, modality-specific regions of cortex are thought to store the perceptual and/or motor experiences associated with a given memory. These regions then form part of a larger network, or memory code, which becomes reactivated during memory retrieval [10]. While we cannot show that this network is automatically activated during memory retrieval (as has been suggested elsewhere, cf. [20]), we do show that the memory code for sensory information is activated early within modality-specific regions of cortex (indeed, during input within the crossmodal channel). What is unique about our paradigm is that the sensory input is matched during our comparison S1 period, so the recruitment of visual cortex during auditory input or auditory cortex during visual input can only be explained by processes involved in the retrieval (i.e., reactivation) of the crossmodal long-term paired associate, and not by sensory-specific processing per se.

Previous fMRI studies have suggested that the learning of crossmodal compared with unimodal paired associates results in increased activity in the associated stimulus modality [3,4], though temporal limitations of fMRI make the precise timing of such activity difficult to interpret. Also, previous fMRI findings suggest that learning results in enhanced activity for unimodal paired stimuli within the corresponding sensory-specific cortex [5], though this finding is difficult to interpret given that there is within-modality sensory-specific input. Here, we clearly add to these previous findings by showing that the learning of long-term paired associates results in activation within the corresponding crossmodal sensory cortex for a given input (i.e., visual cortex for auditory stimuli, auditory cortex for visual stimuli), and that this power difference occurs early (within the S1 period in our study). The timing of this difference is in accord with previous MEG findings which indicate that cortical networks diverge within 500 ms of stimulus onset for crossmodal compared with unimodal paired items [21].

One interpretation of our findings is that greater activation in the associated crossmodal association areas is simply a reflection of greater task difficulty (and therefore greater attentional demands) for crossmodal compared with unimodal stimuli. However, the behavioral results suggest that the greatest attentional demands occur during the unimodal AA condition (both reaction time and error rates suggest this interpretation). Performance during either crossmodal condition (AV and VA) then falls somewhere in between, with the least attentional demands during the unimodal VV condition. Thus, our effects in early sensory cortices cannot be accounted for by attention, as there is no inverse relationship between VA-VV and AV-AA. That is, attention should produce greater activation for VA relative to VV (as VA was more difficult behaviorally) in auditory areas, and also produce less activation in visual areas for AV to AA (as AA was more difficult behaviorally). However, in both comparison conditions we found greater power for the crossmodal relative to unimodal condition within the associated sensory cortex, suggesting that this effect is due to the crossmodal nature of the stimuli independent of task difficulty.

Taken together, our results suggest that retrieval of auditory or visual objects from memory involves the activation of modality-specific cortices. When these memories include crossmodal associations, there is a corresponding co-activation of the associated sensory modality, which occurs in the absence of explicit sensory input. This co-activation is what might ultimately produce facilitated performance for crossmodal compared with unimodal

stimuli. Of course, we recognize that several cognitive processes are involved in the retrieval of paired associates (both crossmodal and unimodal) from memory. Certainly our results demonstrate a host of brain regions showing differences between our crossmodal and unimodal paired items. To fully understand the role of all the involved areas, further sophisticated analyses (e.g. connectivity analysis) will be necessary. However, our findings add further evidence to the idea that memory formation is distributed across modality-specific regions of cortex that store the corresponding perceptual and motor codes related to a given memory [8-11].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the intramural program of the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health. The authors thank Tom Holroyd, Fred Carver, and Jason Smith for technical advice and helpful discussions.

REFERENCES

1. Miyashita, Y. Visual associative long-term memory: Encoding and retrieval in inferotemporal cortex of the primate.. In: Gazzaniga, MS., editor. *The new Cognitive Neurosciences*. 2nd ed. MIT press; Cambridge, MA: 2000. p. 379-92.
2. Miyashita Y, Hayashi T. Neural representation of visual objects: Encoding and top-down activation. *Curr Opin Neurobiol*. 2000; 10:187–94. [PubMed: 10753793]
3. Tanabe HC, Honda M, Sadato N. Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. *J Neurosci*. 2005; 25:6409–18. [PubMed: 16000632]
4. Naumer MJ, Doehrmann O, Muller NG, Muckli L, Kaiser J, Hein G. Cortical plasticity of audio-visual object representations. *Cerebral Cortex*. 2009; 19:1641–53. [PubMed: 19015373]
5. Butler AJ, James KH. Cross-modal versus within-modal recall: Differences in behavioral and brain responses. *Behav Brain Res*. 2011; 224:387–96. [PubMed: 21723328]
6. Bolognini N, Frassinetti F, Serino A, Ladavas E. Acoustical vision of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Exp Brain Res*. 2005; 160:273–82. [PubMed: 15551091]
7. Atkinson RC, Shiffrin RM. The control of short-term memory. *Sci Am*. 1971; 225:82–90. [PubMed: 5089457]
8. Martin A. The representation of object concepts in the brain. *Ann Rev Psychol*. 2007; 58:25–45. [PubMed: 16968210]
9. Barsalou LW. Grounded cognition. *Ann Rev Psychol*. 2008; 59:617–45. [PubMed: 17705682]
10. Fuster JM. Cortex and memory: Emergence of a new paradigm. *J Cog Neurosci*. 2009; 21:2047–72.
11. Meyer K, Damasio A. Convergence and divergence in a neural architecture for recognition and memory. *Trends Neurosci*. 2009; 32:376–82. [PubMed: 19520438]
12. Damasio AR. Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*. 1989; 33:37–43.
13. Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, et al. Activation of auditory cortex during silent lipreading. *Science*. 1997; 276:593–6. [PubMed: 9110978]
14. Gonzalez J, Barros-Loscertales A, Pulvermuller F, Mesequer V, Sanjuan A, Belloch V, Avila C. Reading cinnamon activates olfactory brain regions. *Neuroimage*. 2006; 32:906–12. [PubMed: 16651007]
15. Kiefer M, Sim EJ, Hernberger B, Grothe J, Hoenig K. The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *J Neurosci*. 2008; 28:12224–30. [PubMed: 19020016]

16. Naya Y, Yoshida M, Miyashita Y. Forward processing of long-term associative memory in monkey inferotemporal cortex. *J Neurosci*. 2003; 23:2861–71. [PubMed: 12684473]
17. Fuster JM, Bodner M, Kroger JK. Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*. 2000; 405:347–51. [PubMed: 10830963]
18. Zhou YD, Fuster JM. Visuo-tactile cross-modal associations in cortical somatosensory cells. *Proc Natl Acad Sci USA*. 2000; 97:9777–82. [PubMed: 10944237]
19. Amedi A, Stern WM, Camprodon JA, Bermpohl F, Merabet L, Rotman S, et al. Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat Neurosci*. 2007; 10:687–9. [PubMed: 17515898]
20. Weisberg J, van Turenout M, Martin A. A neural system for learning about object function. *Cereb Cortex*. 2007; 17:513–21. [PubMed: 16581980]
21. Banerjee A, Pillai AS, Sperling JR, Smith JF, Horwitz B. Temporal microstructure of cortical networks (TMCN) underlying task-related differences. *Neuroimage*. 2012; 62:1643–1657. [PubMed: 22728151]
22. Vrba J, Robinson SE. Signal processing in magnetoencephalography. *Methods*. 2001; 25:249–271. [PubMed: 11812209]
23. Tallon-Baudry C. Oscillatory synchrony and human visual cognition. *J Physiol Paris*. 2003; 97:355–63. [PubMed: 14766151]
24. Tallon-Baudry C, Mandon S, Freiwald WA, Kreiter AK. Oscillatory synchrony in the monkey temporal lobe correlates with performance in a visual short-term memory task. *Cereb Cortex*. 2004; 14:713–20. [PubMed: 15054050]
25. Taylor K, Mandon S, Freiwald WA, Kreiter AK. Coherent oscillatory activity in monkey area V4 predicts successful allocation of attention. *Cereb Cortex*. 2005; 15:1424–37. [PubMed: 15659657]
26. Van Veen BD, van Drongelen W, Yucktmann M, Suzuki A. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans Biomed Eng*. 1997; 44:867–880. [PubMed: 9282479]
27. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 1996; 29:162–73. [PubMed: 8812068]

Research Highlights

- Neural correlates of crossmodal associations were explored using MEG
- Crossmodal associations activated corresponding modality-specific sensory cortices
- Crossmodal associations were supported by enhanced beta and gamma power
- These findings suggest memory is distributed across modality-specific cortices

\$watermark-text

\$watermark-text

\$watermark-text

Figure 1. Schematic of stimulus and task design

Panel A shows the four types of pairs used: visual-visual (VV), visual-auditory (VA), auditory-auditory (AA), and auditory-visual (AV). The visual objects were always abstract fractal objects and the sound objects were abstract tone sequences of 1 s duration each. Panel B shows the task design for a single VA trial. During a trial, the first stimulus (S1) was presented for 1 s (here, a visual object), then there was a delay of 1 s (with a fixation cross) followed by a compound stimulus (S2; both an auditory and visual stimulus were presented simultaneously) for 1 s. If either of the compound S2 stimuli included the correct paired associate of S1, the subject responded with a yes button press. Otherwise they were asked to respond with a no button press.

Figure 2. Behavioral Performance

Group-level (N=13) accuracy and reaction time performance for all paired associate types. Error bars indicate standard error of the mean.

Figure 3. Source-Level Findings: Beta Frequency

A) The regions in the visual ventral stream showing increased power in the beta frequency for the contrast AV-AA, including bilateral Cuneus (left and center), and a region in right fusiform gyrus (right) are highlighted by the green circle. B) The region in auditory association cortex (right superior temporal gyrus) showing increased power in the beta frequency for the contrast VA-VV is highlighted by the green circle.

Figure 4. Sliding Window Analyses: Cuneus, Fusiform and STG

A) Sliding window analyses using masks of the regions in Cuneus (left) and fusiform gyrus (right) identified in our beta frequency analysis of AV-AA. Time samples at which the mean power was significantly ($p < 0.05$) greater than zero are denoted by the red overlay. B) Sliding window analysis using a mask of the region in STG identified in our beta frequency analysis of VA-VV. For both conditions, mean power (dark red line) and standard error (dashed red line) across all participants at each time sample are presented within the corresponding ROI (circled in green). No time samples were significantly ($p < 0.05$) greater than zero, but those time samples where p approached significance are noted.

Figure 5. Source-Level Findings: Low Gamma Frequency

A) The region in the visual ventral stream (right middle occipital gyrus) showing increased power in the low gamma frequency for the contrast AV-AA is highlighted by the green circle. B) The region in auditory association (right superior temporal gyrus) cortex showing increased power in the low gamma frequency for the contrast VA-VV is highlighted by the green circle.

Figure 6. Sliding Window Analyses: MOG and STG

A) Sliding window analysis using a mask of the region in MOG identified in our low gamma frequency analysis of AV-AA. Time samples at which the mean power was significantly ($p < 0.05$) greater than zero are denoted by the red overlay. B) Sliding window analysis using a mask of the region in STG identified in our low gamma frequency analysis of VA-VV. For both conditions, mean power (dark red line) and standard error (dashed red line) across all participants at each time sample are presented within the corresponding ROI (circled in green). Time samples at which the mean power was significantly ($p < 0.05$) greater than zero are denoted by the red overlay.

Table 1

Regions Showing Crossmodal-Unimodal Paired Associate Differences: Beta Frequency.

Location	BA	Talairach coordinates (Peak)			Voxels	Pos/Neg	Hemisphere
		X	Y	Z			
AV-AA							
Cuneus	19	-8	-78	33	17	P	L
	19	18	-82	33	5	P	R
Fusiform Gyrus	37	48	-53	-8	5	P	R
Cingulate Gyrus/Caudate		18	-22	27	12	P	R
	25	-2	8	-3	10	P	L
Inferior Frontal Gyrus	11	28	33	-23	11	P	R
Supramarginal Gyrus	40	58	-47	27	9	P	R
Precuneus	19	3	-82	42	8	P	R
Middle Temporal Gyrus/Parahippocampal Gyrus	19	38	-47	2	6	P	R
Middle Frontal Gyrus	6	33	8	57	6	P	R
Superior Temporal Gyrus	5	53	-12	7	5	P	R
Posterior Cingulate	23		-57	17	5	P	R
Precentral Gyrus	6	43	3	37	5	P	R
VA-VV							
Superior Temporal Gyrus	22	63	-7	7	9	P	R
Precentral Gyrus	44	63	8	12	6	P	R
Superior Frontal Gyrus	6		18	52	4	P	R

All regions surviving group-level statistical significance ($p < 0.01$, corrected) in our AV-AA and VA-VV comparisons in the beta (15-30 Hz) frequency. BA=Brodman area; Voxels=number of voxels surviving our significance threshold ($p < 0.01$, corrected); P=positive, N=negative; L=left, R=right.

Table 2
Regions Showing Crossmodal-Unimodal Paired Associate Differences: Low Gamma Frequency.

Location	BA	Talairach coordinates (Peak)			Voxels	Pos/Neg	Hemisphere
		X	Y	Z			
AV-AA							
Middle Occipital Gyrus	37	48	-67	-3	5	P	R
Angular Gyrus	39	48	-62	32	16	P	R
Superior Frontal Gyrus	10	-22	48	27	12	N	L
VA-VV							
Superior Temporal Gyrus	13	58	-32	17	60	P	R
Inferior Frontal Gyrus	47	38	23	-8	49	P	R
Postcentral Gyrus	3	-32	-32	57	38	N	L
Cingulate Gyrus	24	13	-12	32	34	P	R
	31	13	-32	42	31	N	R
Cuneus	19	28	-82	27	34	P	R
	18	3	-72	17	8	N	R
Middle Frontal Gyrus	9	48	18	37	26	P	R
	8	-27	28	47	6	P	L
Inferior Parietal Lobule	13	33	-32	27	12	P	R
Superior Temporal Gyrus (ant./inf.)	38	33	13	-33	8	P	R
Middle Temporal Gyrus	39	48	-72	17	8	P	R
Precuneus	19	38	-72	37	7	P	R
Middle Occipital Gyrus	19	53	-67	-8	6	P	R

All regions surviving group-level statistical significance ($p < 0.01$, corrected) in our AV-AA and VA-VV comparisons in the low gamma (31-54 Hz) frequency. BA=Brodman area; Voxels=number of voxels surviving our significance threshold ($p < 0.01$, corrected); P=positive, N=negative; L=left, R=right.