Auditory and Action Semantic Features Activate Sensory-Specific Perceptual Brain Regions

Thomas W. James* and Isabel Gauthier
Vanderbilt Vision Research Center
Psychology Department
Vanderbilt University
Nashville, Tennessee 37203

Summary

Traditionally, concepts were considered proposi-
tional, amodal, and verbal in nature (for review, see
[1]). Recent findings, however, suggest that concep-
tual knowledge is divisible into different types (L. Wu
and L.W. Barsalou, personal communication, [2]) and
that each type may be linked to specific sensory and
motor processes [3, 4]. This implies that sensory pro-
cessing regions of the brain may also process con-
cepts. In fact, there is some neuroimaging evidence
that conceptual information does activate perceptual
brain regions and that there is a correspondence be-
tween knowledge type and the region being activated
[5, 6]. In the following experiment, using a training
technique developed in previous studies [7, 8], partici-
pants verbally learned associations between novel ob-
jects and conceptual features. The objective was to
create objects that were associated with features from
only one knowledge type, something that does not
occur with common objects. During a visual task that
did not require retrieval of learned associations, the
superior temporal gyrus, which responds well to
sounds, was preferentially activated by objects asso-
ciated with auditory features (e.g., buzzes). Likewise,
the posterior superior temporal sulcus, which re-
sponds well to motion, was preferentially activated by
objects associated with "action" features (e.g., hops).
These findings support the theory that knowledge is
grounded in perception.

Results and Discussion

Training Data

Participants learned to associate three adjectives with
each object in a set of four (see Figure 1 for an example).
Each participant was trained with three sets of novel
objects, using three different sets of adjectives, with
each set of adjectives belonging to a different knowl-
edge type. These three knowledge types and the adjec-
tives themselves were borrowed from a semantic feature
nomenclature that has been used successfully else-
where [4] to describe the variability of recognition defi-
cits seen in patients with category-specific semantic
deficits. Auditory (AUD) features were sounds (e.g.,
buzzes), action (ACT) features were movements per-
formed by another entity (e.g., walks), and encyclopedic
(ENC) features could not be easily assigned to one sen-
sory-specific type (e.g., long lived). Thus, training was

*Correspondence: thomas.james@vanderbilt.edu

used to imbue each of the objects with verbally learned
conceptual knowledge that would be impossible to ap-
prehend based on the static visual appearance of the
object alone. Importantly, this knowledge was always
derived from only one type, something that does not
occur with common objects.

After training (which was accomplished over three
sessions, each approximately 45 min long), learning was
assessed with a criterion test. Participants were given
the triad of features that described an object and made
a three-alternative forced-choice decision from among
the target object and two distractors from the same set.
Accuracy criterion was 15 of 18 correct trials in a block.
All reaction times for correct trials had to be under 4 s.
Participants were given 20 blocks of trials to reach crite-
rion. Of the 13 participants, one did not reach criterion
and was therefore excluded from further analyses.

Behavioral Data

After training, participants performed a sequential
matching task with pairs of objects chosen from the
three trained object sets or from a nontrained (NON) set
of objects. Sensitivity, expressed by a d′ measure, was
high for all conditions (AUD, 5.4 [SEM = 0.34]; ACT, 4.9
[0.44]; ENC, 4.9 [0.42]; NON, 5.2 [0.41]), and a one-way
analysis of variance showed no significant difference
between conditions. Reaction times (ms) were relatively
long (AUD, 3933 [SEM = 462]; ACT, 3634 [404]; ENC,
3358 [328]; NON, 3977 [491]), presumably due to the
perceptual similarity of the objects. Again, a one-way
analysis of variance showed no significant difference
between conditions.

Imaging Data

During imaging, participants matched simultaneously
presented pairs of objects from each of the four condi-
tions in a rapid event-related design paradigm. The
matching task could be performed solely on the basis of
the visual information in the images, thus any differences
between the training conditions should result from inci-
dental retrieval of the trained associations. Objects from
the four conditions were presented in random order.
Objects were presented for 1200 ms, followed by a 2800,
4800, or 6800 ms intertrial interval. Data from 13 partici-
pants were analyzed individually using a region-of-inter-
est (ROI) analysis. We attempted to isolate two ROIs
for each participant, one involved in the processing of
sounds and the other involved in the processing of bio-
logical motion (see Figures 2 and 3 for methods). Five
participants showed reliable activation in the expected
regions for both localizer tasks and four additional parti-
cipants showed reliable activation in only one of the two
tasks. Therefore, the ROI analysis was carried out on
data from these nine participants. Within these ROIs,
separate general linear model (GLM) analyses were car-
ried out using the Brain Voyager™ event-related decon-
volution procedure. This analysis allows for estimation
of the underlying signal for each condition in a ROI,
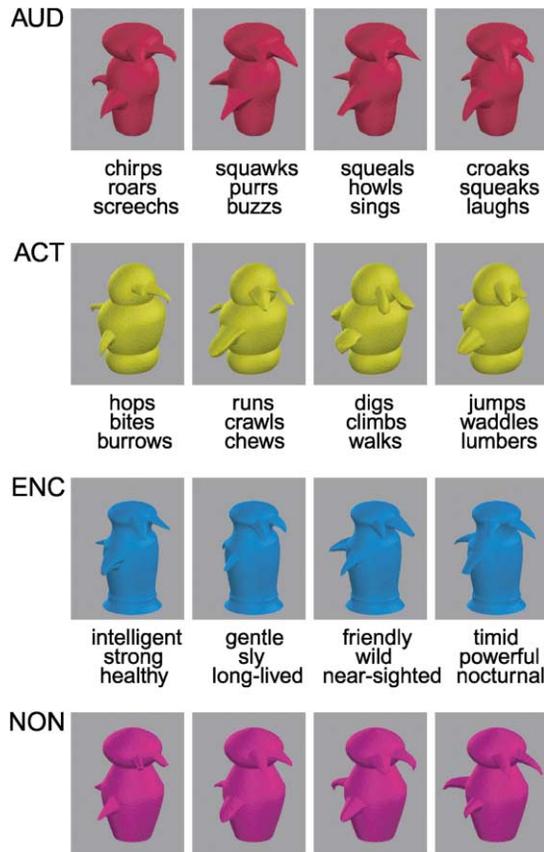
Figure 1. An Example of Novel Object Sets and Semantic Training Conditions

Four sets of four objects (asymmetric "Greebles") each were used (www.cog.brown.edu/~tarr/stimuli). Visual similarity among individual objects was greater within each set than it was between sets. Each set was displayed in a different color to maximize the distinctiveness of each set. In this example, set one was trained with auditory features, set two was trained with action features, set three was trained with encyclopedic features, and set four was not trained.

despite overlap of the hemodynamic response functions for successive rapidly presented trials. Note, however, that deconvolution analyses have poor reliability when used to create statistical parametric maps (R. Marois, personal communication) due to its high susceptibility to noise.

Activated regions from the sound localizer included superior temporal gyrus (STG), superior temporal sulcus (STS), and parts of the frontal and parietal cortices. Because our aim was to localize auditory cortex in the STG, only activated voxels within the STG were included in the STG ROI. Statistical maps showing the STG and other activations produced by the sound localizer are shown in Figure 2 with a plot of the average right STG response for each training condition estimated using event-related deconvolution. ROIs from the motion localizer were found in the posterior superior temporal sulcus (STSp), the middle temporal area (MT+), and another more posterior occipital area. Statistical maps showing the STSp and MT+ are shown in Figure 3 with a plot of the average right STSp response for each training condition estimated using event-related deconvolution.
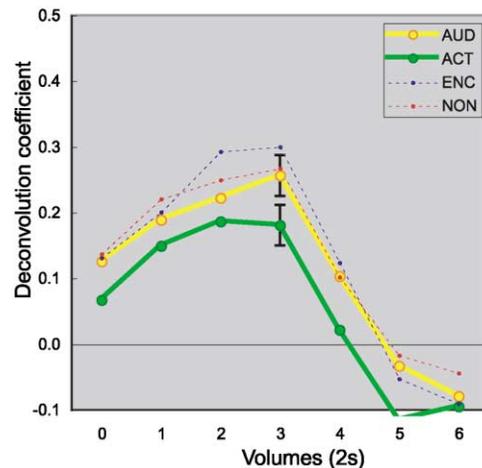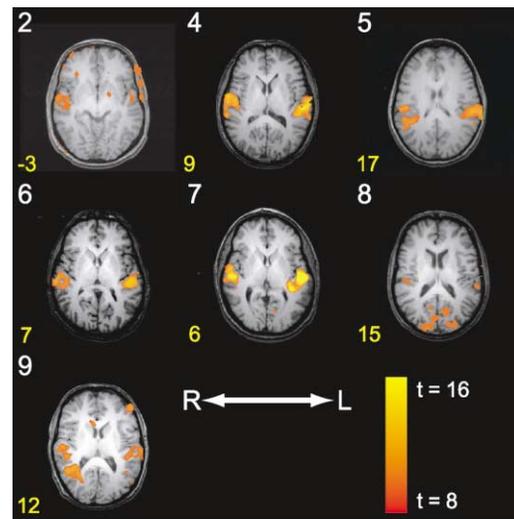
Figure 2. Statistical Maps Comparing Listening to Sounds with the Resting State in Seven Participants

Sounds were all animal noises and were presented in a blocked design paradigm using 16 s blocks of sound interleaved with 10 s blocks of rest. Numbers in the upper left of each brain image identify the participants for comparison with Figure 3. Numbers in the lower right indicate the distance above or below the ACPC plane in mm. The average Talairach coordinates for the superior temporal gyrus region were 52 (SD = 3), −17 (8), and 10 (7). The graph shows time courses derived from the deconvolution analysis for the four training conditions averaged across the seven right-hemisphere STG ROIs. The vertical axis represents the least-squares fit coefficient from the deconvolution analysis, which is related to the theoretical average percent signal change value for that condition. The horizontal axis is time in volumes (2 s per volume).

All training conditions (including NON) produced significant positive-going activations that resembled prototypical hemodynamic functions in the ROIs shown in Figures 2 and 3. Not surprisingly, the normally visually responsive area STSp produced a much greater response to the visually presented objects than did the normally auditorily responsive area STG. In fact, one may have expected based on previous research [9] that visual stimulation should have deactivated the auditory cortex. This model of cortical activation and deactivation, though, is based on the premise that different sen-
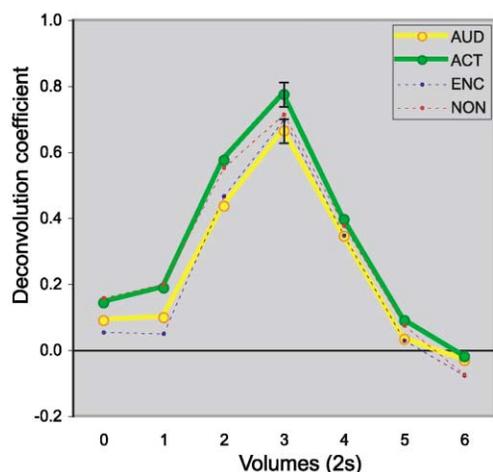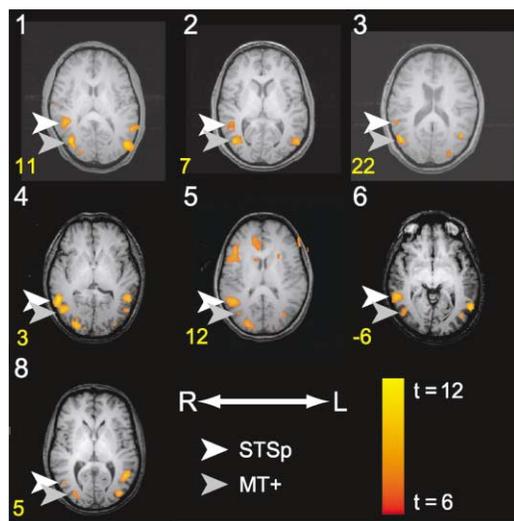
Figure 3. Statistical Maps Comparing Viewing Biological Motion Stimuli with the Resting State in Seven Participants

Stimuli were "point-light" depictions of human movement displayed as black dots on a white background. They were presented in a blocked design paradigm with 16 s blocks of motion interleaved with 10 s blocks of rest. The average Talairach coordinates for the posterior superior temporal sulcus (STSp; indicated by the white arrow) region were 50 (SD = 5), −45 (6), and 9 (7), which falls close to the STSp region described elsewhere [11] as being involved in the processing of biological motion stimuli. The average Talairach coordinates for the middle temporal area (MT+; indicated by the gray arrow) were 41 (5), −66 (5), and 3 (7), which falls close to the MT region described elsewhere [12], a homolog to macaque MT. There were no differences between the training conditions for data taken from the right and left MT+ region. The graph shows time courses derived from the deconvolution analysis for the four training conditions averaged across the seven right-hemisphere STSp ROIs. Axes are as Figure 2.

sory-specific cortices are always competing. Previously, we have shown that associating conceptual features with objects can facilitate visual perceptual judgments of those objects [7], suggesting that in the present paradigm different sensory-specific cortices are *cooperating*. In that case, it may not be unexpected that auditory cortex produced reliable activation to visual stimuli, albeit smaller than that exhibited by truly visual processing

regions. As a precaution, the STG ROI of one participant was divided into sub-ROIs and analyzed further. Although there were small differences between the activation patterns produced by these sub-ROIs, the general pattern of positive-going activation to all object sets was consistent. Therefore, the positive response to visual stimuli seen in our STG ROI was not an artifact of cluster size or location.

Because our a priori hypothesis was that sensory-specific regions would be preferentially activated by objects associated with features belonging to specific knowledge types, a two-way analysis of variance with ROI (STG/STSp) and training condition (AUD/ACT) was performed for each hemisphere; peak deconvolution coefficient was used as the dependent measure for each condition (Figure 4). Despite the fact that five participants produced both ROIs reliably, ROI was treated as a between-groups variable to include the additional four participants that produced only one ROI reliably. This resulted in the analysis of seven independent observations for each ROI. In the right hemisphere, there was a significant interaction between ROI and training condition ($F_{(1,12)}$ = 18.6, $p < 0.001$). Scheffé post hoc tests revealed that AUD objects showed stronger activation than ACT objects for the STG (sound) ROI ($p < 0.05$), and ACT objects showed stronger activation than AUD objects for the STSp (motion) ROI ($p < 0.01$). There was a significant main effect of ROI ($F_{(1,12)}$ = 5.2, $p < 0.05$), with the STSp showing stronger activation during the matching task than the STG. Our event-related task involved visual stimuli only, not auditory stimuli. Thus, it may not be surprising that STSp, a region defined using visual stimuli in this study, responded more strongly than STG, which was defined using auditory stimuli.

Our results reveal that objects associated with semantic features from specific knowledge types preferentially activated sensory-specific regions of cortex. These findings suggest that semantic memory may be stored in sensory/motor modality-specific subsystems instead of a unitary amodal system. In this experiment, the semantic features were not direct sensory impressions but instead were learned verbally. One might speculate that sensory-specific brain regions would respond even more strongly to the associated semantic features if those features were learned through direct sensory experience. Our choice of verbal learning instead of direct sensory experience, however, was motivated by the following reasoning: if semantic memory is amodal, associations with direct sensory stimuli may still activate sensory-specific cortex, but associations with word stimuli should optimally activate only the theoretical amodal store. Therefore, for investigating whether there is a perceptual basis to conceptual knowledge, verbal learning actually provides a stronger test than does direct sensory experience.

Although matching objects associated with auditory and action features produced clearly opposite effects in STG and STSp, the activation produced by the encyclopedic and nontrained objects did not follow an easily interpretable pattern. In STSp, action objects appeared to have special status, producing more activation than the other three conditions. But, in STG, encyclopedic and nontrained objects produced activation equal to
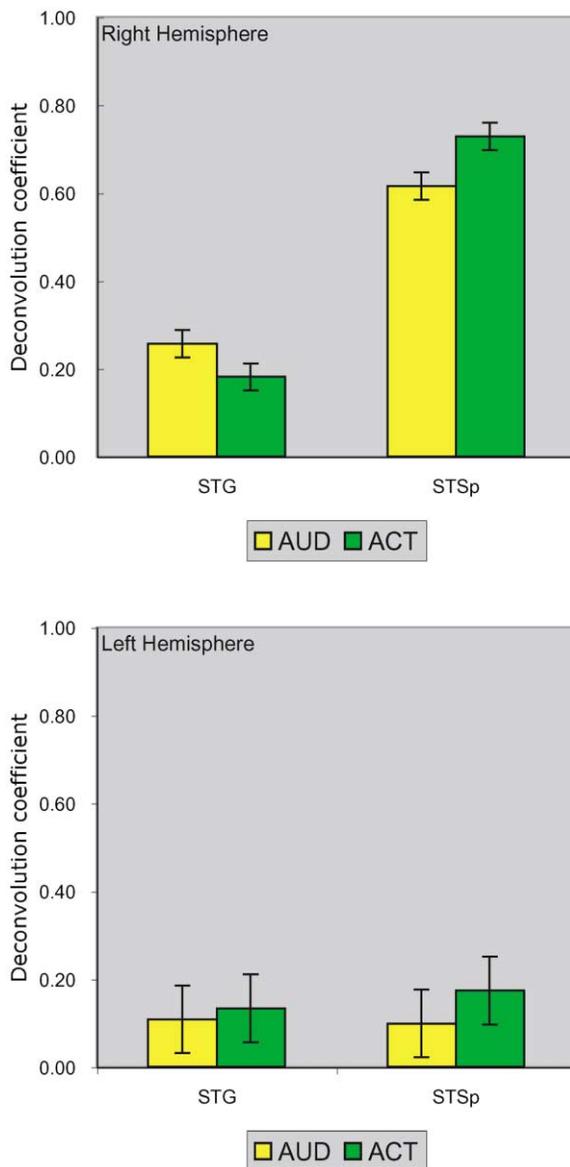
Figure 4. Peak Activation as a Function of Region of Interest and Training Condition

The vertical axis represents the mean across participants of the peak deconvolution coefficient. This coefficient represents the least-squares fit coefficient from the deconvolution analysis. The coefficient is related to the theoretical percent signal change.

that of AUD objects, suggesting that AUD objects were not specially activated but instead that the ACT objects were perhaps specially deactivated. It is possible, even though precautions were taken against it, that some generalization did occur across sets of objects, which may explain the heightened activation of the encyclopedic objects. It is also likely that the nontrained objects showed higher than expected activation because they were less familiar than the trained objects [10].

There were no significant differences between training conditions in ROIs from the left hemisphere. A hemispheric difference was not predicted, but there are at least two possible explanations for why there would be

differences between conditions in the right but not in the left hemisphere. First, the left hemisphere ROIs both showed lower overall activation with the object stimuli than the right hemisphere ROIs. This difference was most marked in STSp. Of note is that the STSp ROI was found more reliably in the right hemisphere than in the left (Figure 3). This suggests that, at least for biological motion, the largest difference between training conditions (and thus between feature types) was found in the region that produced the best response to the modality-specific perceptual stimulus. Thus, the degree of processing of semantic features in sensory-specific brain regions may be linked to the degree of processing of the corresponding perceptual stimulus. An alternative explanation for the lack of differences between conditions in the left hemisphere, however, is that the storage of semantic information in perceptual processing regions is lateralized to the right hemisphere, at least for the two feature types studied here.

Conclusions

Associating novel objects with verbally learned semantic features from different knowledge types produced different patterns of cortical activation during a subsequent perceptual task. An area in the right superior temporal gyrus (STG), which was shown to be involved in the processing of sounds, was preferentially activated by objects associated with auditory features. An area in the right posterior superior temporal sulcus (STSp), which was shown to be involved in the processing of biological motion, was preferentially activated by objects associated with "action" features. Taken together, this dissociation supports the hypothesis that auditory and action knowledge is processed in sensory-specific perceptual processing regions of the brain.

Experimental Procedures

Participants

Participants were graduate students, postdoctoral fellows, or research assistants in the Psychology department at Vanderbilt University. All participants reported that they were right-handed, had normal or corrected-to-normal vision, and had no history of neurological disorders. There were five females and seven males, with ages ranging from 23 to 42 years with a mean age of 29.0. The protocol was approved by the Vanderbilt University Institutional Review Board, and all participants gave informed consent.

Imaging Parameters and Preprocessing

All imaging was done using a 3-Tesla, whole-body GE MRI system and a birdcage head coil located at the Vanderbilt University Medical Center (Nashville, USA). The field of view was $24 \times 24 \times 12.6$ cm, with an in-plane resolution of 64×64 pixels and 18 contiguous oblique axial scan planes per volume (whole brain), resulting in a voxel size of $3.75 \times 3.75 \times 7.0$ mm. Images were collected using a T2*-weighted EPI acquisition (TE, 25 ms; TR, 2000 ms; flip angle, 70°) for blood oxygenation-level dependent (BOLD)-based imaging. High-resolution T1-weighted anatomical volumes were also acquired using a 3D fast spoiled grass (FSPGR) acquisition (TI, 400 ms; TE, 4.18 ms; TR, 10 ms; FA, 20°).

The functional data underwent slice time correction, 3D motion correction, linear trend removal, and gaussian spatial blurring (FWHM 4mm) using the Brain Voyager™ 2D analysis tools. The 3D anatomical volumes, to which the functional volumes were registered, were transformed into the Talairach stereotactic space that was common for all participants.

References

1. Glaser, W.R. (1992). Picture naming. Cognition *42*, 61–105.
2. McRae, K., and Cree, G.S. (2002) Factors underlying category-specific semantic deficits. In Category-Specificity in Mind and Brain. E.M.E. Forde and G. Humphreys, eds. (East Sussex, UK: Psychology Press), in press.
3. Barsalou, L.W., Simmons, W.K., Barbey, A.K., and Wilson, C.D. (2003). Grounding conceptual knowledge in modality-specific systems. Trends Cogn. Sci. *7*, 84–91.
4. Cree, G.S., and McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). J. Exp. Psychol. Gen. *132*, 163–201.
5. Martin, A., Haxby, J.V., Lalonde, F.M., Wiggs, C.L., and Ungerleider, L.G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. Science *270*, 102–105.
6. Kan, I.P., Barsalou, L.W., Solomon, K.O., Minor, J.K., and Thompson-Schill, S.L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. Cogn. Neuropsychol. *20*, 525–540.
7. Gauthier, I., James, T.W., Curby, K.M., and Tarr, M.J. (2003). The influence of conceptual knowledge on visual discrimination. Cogn. Neuropsychol. *20*, 507–523.
8. James, T.W., and Gauthier, I. Brain areas engaged during visual judgements by involuntary access to novel semantic information. Vision Res., in press.
9. Laurienti, P.J., Burdette, J.H., Wallace, M.T., Yen, Y.-F., Field, A.S., and Stein, B.E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. J. Cogn. Neurosci. *14*, 420–429.
10. Schacter, D.L., and Buckner, R.L. (1998). Priming and the brain. Neuron *20*, 185–195.
11. Grossman, E.D., and Blake, R. (2002). Brain areas active during visual perception of biological motion. Neuron *36*, 1167–1175.
12. Dukelow, S.P., Desouza, J.F.X., Culham, J.C., van den Berg, A.V., Menon, R.S., and Vilis, T. (2001). Distinguishing subregions of the human MT+ complex using visual fields and pursuit eye movements. J. Neurophysiol. *86*, 1991–2000.