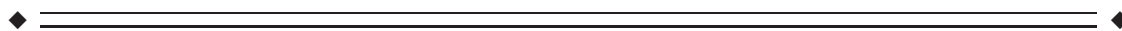


# Representations of Modality-Specific Affective Processing for Visual and Auditory Stimuli Derived From Functional Magnetic Resonance Imaging Data

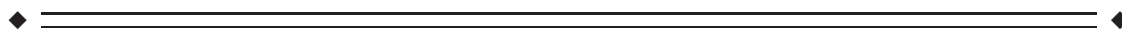
Svetlana V. Shinkareva,\* Jing Wang, Jongwan Kim, Matthew J. Facciani, Laura B. Baucom, and Douglas H. Wedell

*Department of Psychology, University of South Carolina, Columbia, South Carolina*



**Abstract:** There is converging evidence that people rapidly and automatically encode affective dimensions of objects, events, and environments that they encounter in the normal course of their daily routines. An important research question is whether affective representations differ with sensory modality. This research examined the nature of the dependency of affect and sensory modality at a whole-brain level of analysis in an incidental affective processing paradigm. Participants were presented with picture and sound stimuli that differed in positive or negative valence in an event-related functional magnetic resonance imaging experiment. Global statistical tests, applied at a level of the individual, demonstrated significant sensitivity to valence within modality, but not valence across modalities. Modality-general and modality-specific valence hypotheses predict distinctly different multidimensional patterns of the stimulus conditions. Examination of lower dimensional representation of the data demonstrated separable dimensions for valence processing within each modality. These results provide support for modality-specific valence processing in an incidental affective processing paradigm at a whole-brain level of analysis. Future research should further investigate how stimulus-specific emotional decoding may be mediated by the physical properties of the stimuli. *Hum Brain Mapp* 35:3558–3568, 2014. © 2013 Wiley Periodicals, Inc.

**Key words:** valence; modality; fMRI; multivariate pattern analysis; STATIS



## INTRODUCTION

The theory of core affect posits that the neural system processes affective aspects of stimuli encountered by the organism quickly and automatically, resulting in a unified affective state described along the dimensions of valence and arousal (Russell, 2003; Russell and Barrett, 1999). Studies of core affect have typically presented visual or auditory stimuli scaled for affective content to participants, who may then evaluate them along affective dimensions or may process them incidentally (e.g., Lang et al., 1998). Because core affect is typically studied as a reaction to stimulus presentation, researchers have posited that there

---

Contract grant sponsors: The Social Sciences Grant program, Office of the Executive Vice President for Academic Affairs and Provost, University of South Carolina.

\*Correspondence to: Svetlana V. Shinkareva, Department of Psychology, University of South Carolina, Columbia, South Carolina 29208, USA. E-mail: shinkareva@sc.edu

Received for publication 7 June 2013; Revised 7 October 2013; Accepted 9 October 2013.

DOI 10.1002/hbm.22421

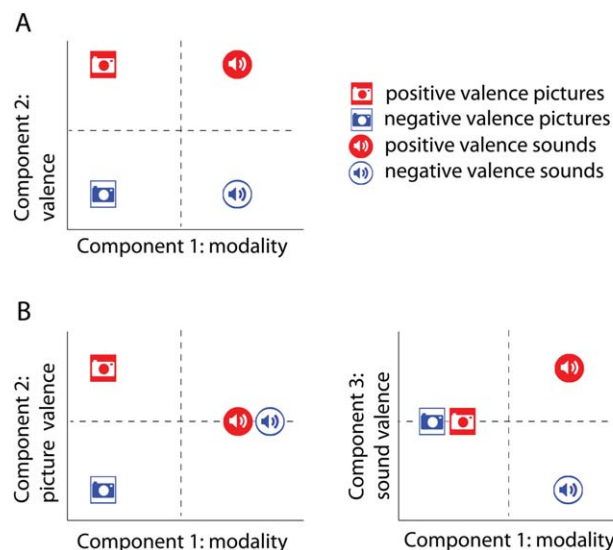
Published online 2 December 2013 in Wiley Online Library (wileyonlinelibrary.com).

may be two functional systems involved in processing the affective content of the stimuli (Barrett and Bliss-Moreau, 2009). The first system is conceived as a sensory-integration network that is modality-specific, so that visual stimuli are processed in a different network than auditory stimuli. Thus, affective processing within the sensory-integration network would likely be modality-specific, with different neural encoding of affective sounds than of affective pictures. A second functional subsystem for core affect processing is the visceromotor network (Barrett and Bliss-Moreau, 2009). In this network an affective response is generated to the stimulus, with this response likely to be modality-general so that negatively valenced sounds may be processed in a similar way as negatively valenced pictures.

The study presented here explores how modality-specific and modality-general encoding may be distinguished using functional magnetic resonance imaging (fMRI) methodology in three different ways. All three approaches are based on a multivariate pattern analysis framework in which distributed patterns of activation across multiple voxels contribute to the representation of the affective state. The first approach is based on decoding affective states. If modality-general processing is occurring, then one would expect that not only should affect be decoded within a given modality but also that cross-modality decoding of affect should occur. If only modality-specific processing of affect is occurring, then decoding should be successful within modality, but not across modalities.<sup>1</sup>

The second approach is an analysis of the variance associated with different contrasts applied to each voxel. If modality-general affective processing is occurring, then one would expect the variance of a global affective contrast that ignores modality differences to be significantly greater than chance. If modality-specific processing dominates, then modality-specific affective contrasts should be significant and modality-general contrasts should not.

The third approach does not utilize significance testing, but rather visualization methods. Possible representations that might arise when modeling similarity of neural responses to positively and negatively valenced visual or auditory stimuli are illustrated in Figure 1. Panel A portrays the modality-general valence hypothesis, with principal components representing modality and valence, respectively, and stimulus locations indicating a common valence component that functions in the same way for both modalities. Panel B portrays a three-component representation of how these stimuli might be affectively represented if valence is encoded in different ways for pictures and sounds. In the three-component representation, it is clear that valence for pictures does not predict valence for sounds, and vice versa, with modality-specific components for coding picture valence and sound valence being independent of one another. The representations of Figure 1



**Figure 1.**

Theory-based representations of the affective space for positive and negative valence generated by pictures and sounds. **(A)** Modality-general valence in a two-dimensional space. **(B)** Modality-specific valence in a three-dimensional space. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

suggest that applying multidimensional scaling related methods of voxel similarity patterns (Baucom et al., 2012; Edelman et al., 1998; Kriegeskorte et al., 2008a, 2008b; Shinkareva et al., 2013) across modality-by-valence conditions may be useful in examining the nature of the dependency of affect and modality.

Following the logic outlined above, we utilized an experimental design in which we manipulated stimuli along two dimensions: modality (visual or auditory) and valence (positive or negative). The multivariate patterns of brain activity associated with experimental conditions have been previously shown to contain information that is not available in activation-based analyses (e.g., Jimura and Poldrack, 2012) and have been successfully analyzed in the study of affect (Baucom et al., 2012; Ethofer et al., 2009; Kassam et al., 2013; Kotz et al., 2013; Peelen et al., 2010; Pessoa and Padmala, 2007; Said et al., 2010; Sitaram et al., 2011; Yuen et al., 2012). These patterns can be compared across conditions to investigate the representational similarity of affective states (Shinkareva et al., 2013). Our aim was to test for modality-specific and modality-general processing in an incidental presentation paradigm in which participants simply viewed or listened to the presented stimuli.

## METHODS

### Participants

Eight (four female) volunteer adults [mean age 27.75 years, standard deviation (SD) = 6.41] from the University

<sup>1</sup>Lack of cross-modal decoding, however, does not in itself demonstrate modality-specific processing.

of South Carolina community with normal or corrected to normal vision participated and gave written informed consent in accordance with the Institutional Review Board at the University of South Carolina.

### Materials

Participants viewed affect-eliciting pairs of pictures and listened to affect-eliciting pairs of sounds that reflected positive or negative valence. A series of color photographs (e.g., baby smiling) and a series of complex (vocal and nonvocal) naturally occurring sounds (e.g., baby cooing) were sampled from the International Affective Picture System (IAPS) and International Affective Digitized Sounds (IADS) databases, respectively (Bradley and Lang, 2000; Lang et al., 2008), provided by the National Institute of Mental Health Center for the Study of Emotion and Attention. Sounds and pictures were selected based on normed valence ratings and balanced across the two valence categories on arousal ratings (Table I). Note that the positive pictures differ strongly from the negative pictures in mean valence ratings ( $M_{\text{Positive}} = 7.37$  and  $M_{\text{Negative}} = 2.24$ ) and that the positive sounds differ strongly from the negative sounds in mean valence ratings ( $M_{\text{Positive}} = 7.25$  and  $M_{\text{Negative}} = 2.34$ ). Pictures and sounds were equated on the presence of people and human voices across the two valence categories. On each trial a pair of sounds or pictures was presented so that a given trial did not have a single semantic identity.

We did not attempt to equate the selected stimuli on low-level parameters, although we tested for significant differences in low-level parameters of picture and sound stimuli. First, we measured mean saturation and brightness of pictures using MATLAB (R2010b, MathWorks) and found no significant difference in mean saturation between positive and negative pictures, whereas mean brightness was significantly different between the two conditions,

$t(22) = 2.510$ ,  $P < 0.05$ , with positive pictures brighter than negative pictures ( $M_{\text{Positive}} = 0.601$  and  $M_{\text{Negative}} = 0.470$ ). This relationship is representative of the database itself (Lakens et al., 2013) based on an investigation of all IAPS and Geneva Affective Picture Database (Dan-Glauser and Scherer, 2011) pictures. Baucom et al. (2012) normalized saturation and intensity (brightness) of IAPS picture stimuli and were still able to identify affective values of pictures from fMRI data. Thus brain differences associated with valence appear to be driven by activation differences beyond those in low level vision. We also measured acoustical features of sound stimuli including mean frequency and amplitude between positive and negative sounds using Adobe Audition CS6. Positive sets did not differ from the negative sets in mean frequency (Hz), and mean amplitude (dB),  $P > 0.05$ . We note that while these stimulus sets do not differ significantly on a number of lower level features, this is not an exhaustive search of all physical characteristics of the stimuli.

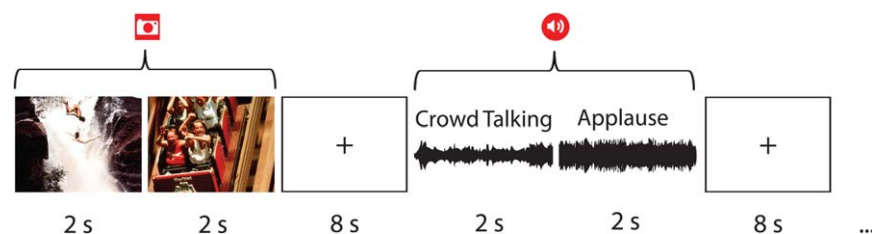
### Experimental Paradigm

fMRI was used to measure brain activity while participants viewed pairs of color photographs and listened to pairs of sounds designed to elicit affective states with either positive or negative valence. All stimuli were presented using E-prime software (Psychology Software Tools, Sharpsburg, PA). All picture stimuli were  $640 \times 480$  pixels and were presented in 32-bit color. Sound stimuli were delivered via Serene Sound Audio System (Resonance Technology, Northridge, CA). Each pair of picture or sound stimuli was presented for 4 s, followed by an 8 s fixation. In each case, two stimuli of the same valence and modality were presented back to back for 2 s to comprise the 4 s presentation (Fig. 2). There were four presentations of 24 unique exemplars, 12 picture pairs and 12 sound pairs, in the experiment. Participants were instructed to

**TABLE I. Picture and sound stimulus pairings**

	Stimulus sets			
	Pictures		Sounds	
	Positive valence	Negative valence	Positive valence	Negative valence
Identification numbers	8170, 8180	9250, 9050	717, 311	260, 292
	8190, 5626	9921, 6821	716, 367	711, 626
	5470, 8034	8485, 9252	351, 813	288, 290
	8499, 8490	2811, 3500	364, 226	424, 625
	5629, 8030	9635, 6312	352, 363	276, 286
	8080, 8200	9600, 9410	817, 815	714, 719
Valence	$M = 7.37$	$M = 2.24$	$M = 7.25$	$M = 2.34$
	$SD = 0.38$	$SD = 0.34$	$SD = 0.77$	$SD = 0.63$
Arousal	$M = 6.42$	$M = 6.60$	$M = 6.39$	$M = 7.25$
	$SD = 0.37$	$SD = 0.25$	$SD = 0.63$	$SD = 0.47$

Note: Statistics summarize nine-point IAPS and IADS ratings of pictures and sounds in each set.



**Figure 2.**

A schematic representation of the presentation timing showing two positive valence trials, picture and sound. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

focus on the fixation cross in the center of the screen throughout the experiment. The presentation sequence was block randomized with the restriction that no affect-by-modality condition was presented twice in a row.

After the scanner session, participants completed a behavioral task in which each of the 12 picture pairs and 12 sound pairs was presented 10 times in random order. After each presentation, participants rated on a 9-point scale their affective reaction to the stimuli on one of the following dimensions: angry, anxious, calm, disgusted, excited, happy, negative, positive, relaxed and sad. Correlations of the pattern of ratings across the 24 stimuli were then used to scale affective reactions to the stimuli.

### fMRI Procedure

MRI data were acquired on a Siemens Magnetom Trio 3.0T whole-body scanner (Siemens, Erlangen, Germany) at the McCausland Center for Brain Imaging at the University of South Carolina. The functional images were acquired using a single-shot echo-planar imaging pulse sequence [repetition time (TR) = 2,200 ms, echo time (TE) = 35 ms, and 90° flip angle] with a 12-channel head coil. Thirty-six 3-mm thick oblique-axial slices were imaged in interleaved scanning order with no gap. The acquisition matrix was 64 × 64 with 3 mm × 3 mm × 3 mm voxels. Functional data was acquired using slow event-related design in a single session. High-resolution whole-brain anatomical images were acquired using a standard T1-weighted 3D MP-RAGE protocol (TR = 2,250 ms, TE = 4.18 ms, field of view = 256 mm, flip angle = 9°, and voxel size = 1 mm × 1 mm × 1 mm) to facilitate normalization of the functional data.

### fMRI Data Processing and Analysis

Data processing and statistical analyses were performed in MATLAB environment using standard procedures in Statistical Parametric Mapping software (SPM 8; Wellcome Department of Cognitive Neurology, London, UK). The data were corrected for motion and linear trend. Structural data were segmented into white and gray matter to facilitate the normalization. Functional and anatomical images were coregis-

tered and spatially normalized into the standard Montreal Neurological Institute space based on T1-derived normalization parameters. Before statistical tests and lower dimensional representation analyses the data were smoothed with an 8 mm full width at half maximum Gaussian kernel.

A general linear model (GLM) was fit at each voxel by convolving the canonical hemodynamic response function with onsets for the conditions of interest, including six motion parameters as nuisance regressors. Low-frequency noise was removed by a filter with a cutoff of 128 s and serial correlations were taken into account using an autocorrelation model AR(1). For multivoxel pattern analysis, a regressor for each of the 96 conditions (12 exemplars × 2 modalities × 4 presentations) was entered into the GLM. Patterns of activation were formed by beta estimates for each of 96 conditions scaled by the square root of the estimated residual variance for each voxel (Misaki et al., 2010). Furthermore, the data for each condition were standardized across voxels to have zero mean and unit variance (Pereira et al., 2009). For statistical tests 24 betas, one for each of the 24 unique exemplars, were estimated based on four presentations each. Finally, for lower dimensional representation analyses eight betas were estimated in total, one for odd and one for even trials for each modality-by-valence combination so that these eight exemplars did not share stimuli in common. We have used multiple trials to obtain more stable beta estimates (Connolly et al., 2012) and to further diminish the effects of semantic identity for this analysis. The eight beta maps were scaled by the square root of the estimated residual variance for each voxel (Misaki et al., 2010).

### Multivariate Pattern Analyses

The multivariate pattern analysis methods employed in this work are similar to those that have been successfully used in our other studies (Baucom et al., 2012; Shinkareva et al., 2011; Wang et al., 2013). Multinomial logistic regression and logistic regression classifiers (Bishop, 2006) were trained to identify patterns of brain activity associated with affect-eliciting visual and auditory stimuli. Four-category classification was performed to identify one of the four conditions associated with affect-eliciting stimuli:



visual positive-valence trials (VP), auditory positive-valence trials (AP), visual negative-valence trials (VN), and auditory negative-valence trials (AN). In addition, two-way classifications were performed to identify positive-valence or negative-valence trials within visual and auditory modality separately, as well as positive-valence or negative-valence trials across the two modalities.

Classification performance was evaluated with fourfold cross-validation, where each fold corresponded to one presentation of the 24 exemplars. Before classification, trials were divided into training and test sets. To reduce the size of and the noise in the data, we performed feature (voxel) selection on the training set only. For each cross-validation fold, a mean of standardized and scaled beta values was calculated across the three replications for each exemplar. A voxel was selected if the absolute value of a mean score surpassed a given threshold for at least one of the 24 exemplars. The classifier was constructed using the selected features from the training set and was applied subsequently to the unused test set. Classification accuracies were computed based on the average classification accuracy across the four cross-validation folds. Statistical significance for the classification accuracies was evaluated by comparison to an empirically derived null distribution constructed by 2,000 noninformative permutations of labels in the training set.

We report the main classification results for multiple threshold levels, rather than deciding upon an arbitrary threshold. For conciseness, confusion matrices and two-way classification accuracies are reported for a single, representative threshold level.

### Testing Modality-Specific and Modality-General Hypotheses

One beta was estimated for each of the 24 unique exemplars based on the four replications. To provide a statistical test of modality-specific hypothesis for each individual, we compared the SD of contrasts reflected in the 2 (modality)  $\times$  2 (valence) design across all gray matter voxels (adjusted probability  $\geq 0.6$  based on SPM segmentation) to contrasts that were orthogonal to the design contrasts. We generated five sets of design-related contrasts to test sensitivity to: (1) visual versus auditory trials, (2) positive versus negative trials, (3) visual-positive versus visual-negative trials, (4) auditory-positive versus auditory-negative trials, and (5) the two-way interaction (visual-positive and auditory-negative versus visual-negative and auditory-positive trials). The sampling distribution of SDs under the null hypothesis was modeled by generating 10,000 contrasts that were orthogonal to the design contrasts.

### Lower Dimensional Representation

STATIS (Lavit et al., 1994), a generalization of principal components analysis for multiple matrices, was used to

investigate the lower dimensional representation of affective space from the functional pattern of whole-brain activity elicited by viewing pictures and listening to sounds from each of the two valence categories. This technique has been previously used in neuroimaging literature (Abdi et al., 2009; Churchill et al., 2012; Kherif et al., 2003; O'Toole et al., 2007; Shinkareva et al., 2008, 2012) and is reviewed in great detail in Abdi et al. (2012). A single  $8 \times 8$  condition-by-condition cross-product matrix was constructed for each participant and these individual cross-product matrices were then analyzed by STATIS. Briefly, a compromise matrix, representing the agreement across individual cross-product matrices, was computed as a weighted average of individual cross-product matrices. Participant weights were derived from the first principal component of the participant-by-participant similarity matrix. Pairwise participant similarity was evaluated by the RV-coefficient (Robert and Escoufier, 1976), a multivariate generalization of the Pearson correlation coefficient to matrices. A compromise matrix was further analyzed by principal components analysis.

We examined the lower dimensional representations derived from the different subsets of voxels using feature selection described above. Crucially, the feature selection method focused on excluding voxels that were not related to any of the exemplars and was not informed about modality or valence contrasts. We describe the behavior of STATIS results for multiple threshold levels and present the results for a single, representative threshold level, for conciseness.

### Analysis of Behavioral Ratings

The correlation matrix generated for the 24 exemplars for each participant was submitted to multidimensional scaling. Scale values from the first dimension were used as an index of the positive or negative reaction to each stimulus pair. Scale values were then evaluated using a 2 (modality)  $\times$  2 (valence)  $\times$  6 (exemplar) repeated measures analysis of variance.

## RESULTS

### Behavioral Ratings

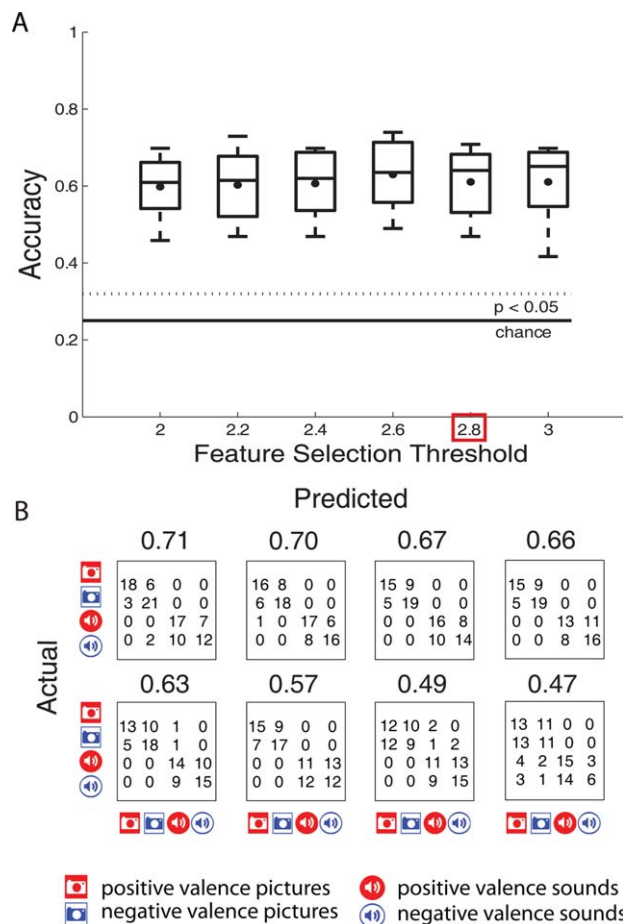
Scale values from the first dimension of multidimensional scaling solutions of behavioral ratings for each participant were submitted to a 2 (modality)  $\times$  2 (valence)  $\times$  6 (exemplar) repeated measures analysis of variance. Only two effects were statistically significant. First, there was a main effect of valence,  $F(1, 7) = 140.9$ ,  $P < 0.001$ , with the mean for positive pairs (0.87) significantly greater than for negative pairs ( $-0.87$ ). Second, there was a significant valence-by-modality interaction,  $F(1, 7) = 19.5$ ,  $P < 0.01$ , in which the difference in valence values was greater for visual stimuli ( $M_{\text{Positive}} = 0.94$  and  $M_{\text{Negative}} = -1.03$ ) than

for auditory stimuli ( $M_{\text{Positive}} = 0.79$  and  $M_{\text{Negative}} = -0.71$ ). The reduced valence differences for sounds may be due to ambiguity in identifying the 2 s samples of the sounds. Nevertheless, the valence differences for both modalities were highly reliable, as revealed by simple effect contrasts,  $F(1, 7) = 229.3$ , partial  $\eta^2 = 0.970$  for pictures and  $F(1,7) = 72.0$  for sounds, partial  $\eta^2 = 0.911$ ,  $P_s < 0.001$ .

### Identification of Affective States

We examined whether the patterns of activity associated with affect-eliciting visual and auditory stimuli had identifiable neural signatures. First, a classifier was trained for each participant to determine if patterns of activity associated with affect-eliciting pictures and sounds were identifiable based on whole-brain activation patterns. For all participants, accuracies for classification of the four conditions (VP, VN, AP, and AN) significantly ( $P < 0.05$ ) exceeded the chance level (0.25) for all the examined levels of thresholds (Fig. 3A). Examination of confusion matrices revealed that although most errors were made within the two modalities, the above chance classification accuracies could not be solely attributed to the modality. For conciseness, we present further results for a single feature selection cutoff, 2.8. The confusion matrices along with classification accuracies for each of the participants are shown in Figure 3B. The highest classification accuracy that was achieved for any of the participants was 0.71, compared to 0.25 chance level. Second, we attempted to identify valence within each of the modalities separately. We were able to identify valence for five of the eight participants in either modality ( $P < 0.05$ ). These are the same participants that showed clear diagonal pattern in the confusion matrices for four-way classification (Fig. 3B). The highest classification accuracy that was achieved for valence identification was 0.81 for pictures and 0.73 for sounds, compared to 0.50 chance level. Third, we attempted to identify valence across the two modalities (i.e., training on valence elicited by sounds data and predicting valence elicited by pictures). These classification accuracies were at chance level.

In auxiliary analyses we investigated whether information in primary sensory cortices is required for valence identification. First, we have attempted to identify valence in picture or sound data based solely on information in the corresponding primary sensory cortices (BA 17 for visual and BA 41 and 42 for auditory stimuli). These analyses did not use feature selection. Above-chance accuracies were found in four of the eight participants for picture data and in three of the eight participants for sound data. Second, we trained the classifiers to identify valence in picture or sound data excluding the corresponding primary sensory cortices. We were able to identify valence in pictures for five of the eight participants and valence in sounds for four of the eight participants. Taken together, these results suggest that there are significant pattern differences, not solely attributable to low-level parameters of



**Figure 3.**

(A) Four-way classification accuracies across the 8 participants, summarized by box plots, are shown for different feature selection thresholds. (B) Confusion matrices and the corresponding classification accuracies for the 8 participants, ordered by classification accuracies, are shown for 2.8 feature selection threshold. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

the stimuli, among the two valence categories within each of the modalities and, therefore, the internal representation of affect could be examined further.

### Testing Modality-Specific and Modality-General Hypotheses

To examine whether the data for each participant are consistent with modality-specific or modality-general hypotheses, we compared the SD of contrasts reflected in the 2 (modality)  $\times$  2 (valence) design across all gray matter voxels to contrasts that were orthogonal to the design contrasts. All eight participants exhibited significant ( $P < 0.01$ ) contrasts for modality. Seven of the eight participants exhibited significant ( $P < 0.01$ ) contrasts for valence within

**TABLE II. Standard deviations for design contrasts, along with  $P$  values, shown in parentheses, based on 10,000 orthogonal contrasts**

Participant	V – A	VP – VN	AP – AN	P – N	VP – VN – AP + AN
1	72.20 (0.00)	30.83 (0.00)	26.03 (0.06)	17.98 (0.97)	22.15 (0.41)
2	84.21 (0.00)	25.03 (0.17)	28.78 (0.01)	17.86 (0.99)	20.21 (0.90)
3	81.52 (0.00)	31.82 (0.00)	29.41 (0.00)	20.79 (0.90)	22.50 (0.51)
4	66.58 (0.00)	27.15 (0.00)	24.59 (0.00)	18.46 (0.69)	18.17 (0.75)
5	48.14 (0.00)	31.03 (0.00)	32.14 (0.00)	21.29 (0.91)	23.33 (0.55)
6	53.45 (0.00)	26.39 (0.00)	25.65 (0.00)	17.80 (0.69)	18.98 (0.39)
7	50.62 (0.00)	33.80 (0.00)	36.06 (0.00)	23.91 (0.94)	25.49 (0.77)
8	50.99 (0.00)	27.11 (0.00)	29.45 (0.00)	19.17 (0.99)	20.86 (0.86)

Participants are ordered by four-way classification accuracies.

Note: To evaluate significance at  $\alpha = 0.05$ , correcting for multiple comparisons,  $P$  values, shown in parentheses, can be compared to 0.01.

visual modality and seven of the eight participants exhibited significant ( $P < 0.01$ ) valence within auditory modality (Table II). None of the participants showed a significant contrast for valence-general modality or interaction contrast ( $P > 0.01$ ). The upshot of this analysis is that while there is clear evidence for significant variation attributable to modality-specific valence variation, there is little evidence supporting significant modality-general valence variation at a whole-brain level of analysis.

### Lower Dimensional Representation of Affective Space

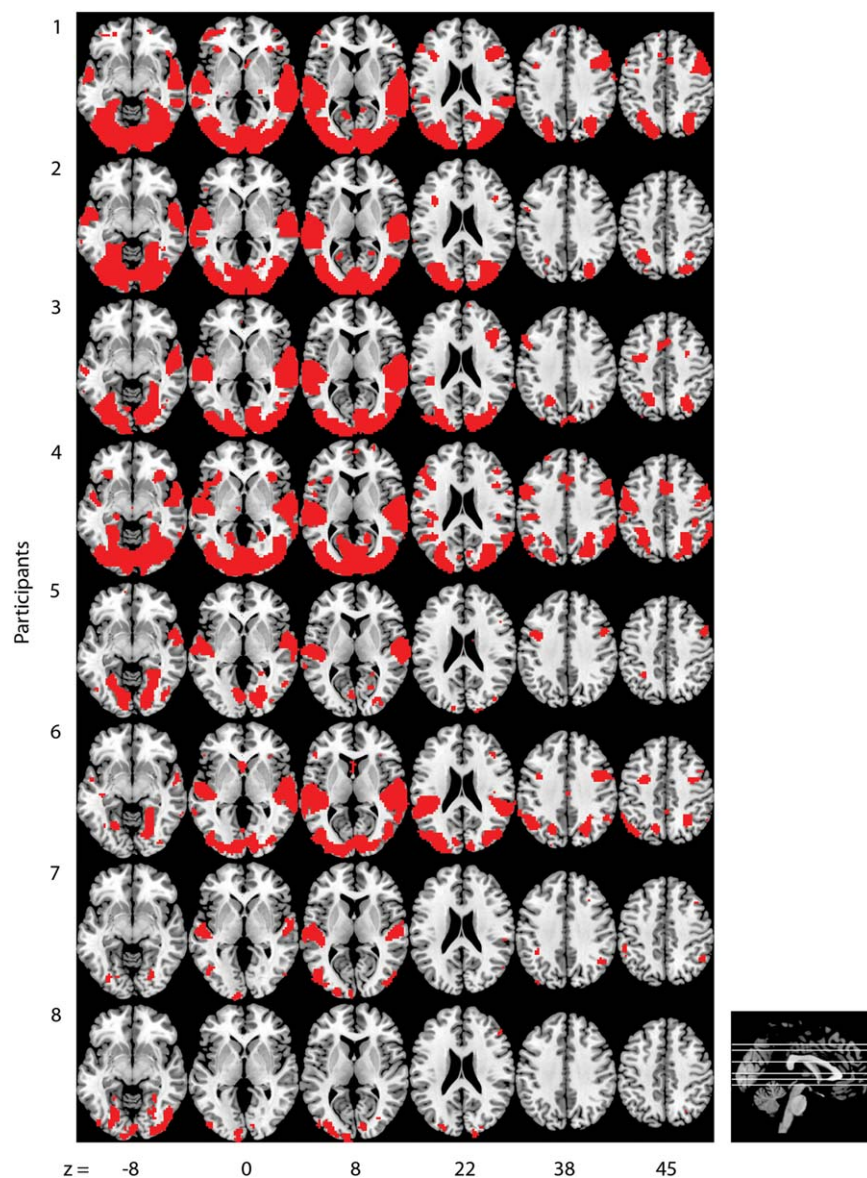
An initial STATIS analysis was conducted using all the gray matter voxels for each participant. The compromise matrix explained 86.61% of the variability in the individual cross-product matrices, thus warranting its interpretation. The first component in this analysis explained 70.12% of the total variance and clearly separated between visual and auditory conditions. The second component explained 5.98% of the total variance and clearly separated positive and negative visual conditions. The rest of the components were not easily interpretable. The lack of valence separation for auditory condition was likely due to inclusion of a large number of noisy voxels. A subsequent STATIS analysis was conducted with a subset of voxels that were sensitive to at least one of the eight exemplars (Fig. 4). This feature selection was not informed about the modality or valence. The compromise matrix explained 94.96% of the variability in the individual cross-product matrices, thus warranting its interpretation. The examination of the optimal set of weights for the eight participants revealed that participants were generally similar in their representation of affective states. Participants 8 and 7 had the smallest weights, and thus contributed less to the compromise. Note that these are the same participants that had the lowest four-way classification accuracies (Fig. 3B) as well as nonsignificant within-modality identifications of valence. Examination of the lower dimensional representation revealed that a three-dimensional space provided the most interpretable solution. Together the three

dimensions explained 96.42% of the total variance and are easily interpretable. The first component explained 92.76% of the total variance and represented modality, with strong separation between visual and auditory conditions. The second component explained 2.23% of the total variance and clearly separated positive and negative visual conditions, whereas the third component explained 1.43% of the total variance and clearly separated positive and negative auditory conditions (Fig. 5). This pattern of results is consistent with the predicted pattern from the modality-specific processing hypothesis shown in Figure 1. Taken together, the classification, statistical tests, and lower dimensional representation results provide evidence for modality-specific valence processing and little indication of modality-general valence processing at the whole-brain level of analysis.

## DISCUSSION

Core affect theory proposes two relevant functional neural networks for processing affective stimuli, the sensory integration network and the visceromotor network (Barrett and Bliss-Moreau, 2009). Having two distinct networks supports the possibility of modality-specific valence processing, although it does not require it. Our presentation of affective stimuli using either auditory or visual modalities provided an opportunity to examine within the fMRI signal the relationship between valence processing and sensory modality. The modality-specific valence hypothesis predicted that there are voxels sensitive to valence differences in one modality that are insensitive to valence differences in the other modality. The modality-general valence hypothesis predicted that there are voxels sensitive to valence differences across different modalities. Our results provided several lines of evidence that supported modality-specific valence processing, but no lines of evidence supporting modality-general valence processing.

Decoding results demonstrated that for a majority of participants valence within modality could be significantly predicted but for no participants was there significant prediction across modalities. Global tests for each individual



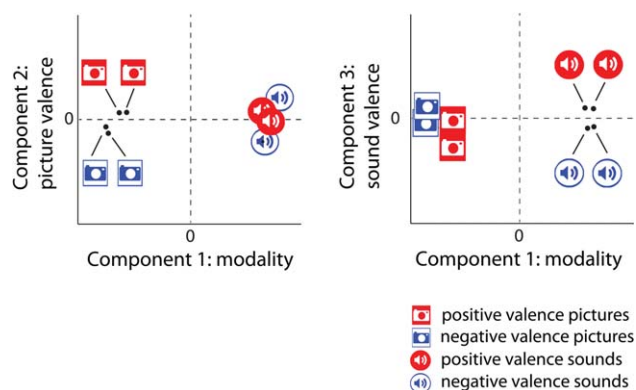
**Figure 4.**

Locations of voxels used for STATIS analysis, whose average magnitude of scaled response surpassed six for at least one of the eight exemplars, are shown for each participant. Participants are ordered by the four-way classification accuracies. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

of contrasts applied to the full set of gray matter voxels demonstrated statistically significant sensitivity of voxels to modality, valence within the visual modality, and valence within the auditory modality, but not for valence across the two modalities. Lower dimensional representations of similarity patterns across voxels supported modality-specific processing. Valence differences for pictures were orthogonal to valence differences for sounds (Fig. 5). Thus, the pattern of results provides support for modality-

specific valence processing and no support for modality-general valence processing at the whole-brain level of analysis. Note that because we did not manipulate arousal, it is unclear whether arousal differences associated with auditory and visual stimuli are encoded in a modality-specific or modality-general way. In previous work with pictures (Baucom et al., 2012), multidimensional scaling of voxel activation patterns produced a circumplex pattern in which valence and arousal were independent dimensions. Because





**Figure 5.**

Positive and negative valence for picture and sound stimuli in the space defined by the first three principal components of the compromise matrix. Component one accounts for 92.76% of the variability in the data and contrasts visual and auditory stimuli. Component two accounts for 2.22% of the variability in the data and contrasts positive and negative valence for picture stimuli. Component three accounts for 1.43% of the variability in the data and contrasts positive and negative valence for sound stimuli. Lines are used to label points in close proximity. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

of this independence, modality-specific valence processing found in the current study has no implications for how arousal is processed. Future research can address this issue.

The finding of modality-specific valence processing is consistent with studies indicating that affective stimuli can enhance processing in sensory-relevant cortices. Valence has been shown to influence sensory processing of both visual (e.g., Lang et al., 1998; Mourao-Miranda et al., 2003) and auditory stimuli (e.g., Fecteau et al., 2007; Plichta et al., 2011; Viinikainen et al., 2012). Higher level sensory processing has also been implicated in processing of emotion stimuli. For example, Said et al. (2010) demonstrated that regions of the superior temporal sulcus can be used to decode facial expressions. Although these previous studies support sensory-specific affective processing, our results differ in at least two respects. First, our study tested for modality-specific affective processing by including affective stimuli across modalities so that we could test for valence relationships between auditory and visual trials. The previous studies examined a single type of stimulus modality and tested in a specific region of interest linked to sensory processing of the stimuli. Second, we examined differential processing of positive and negative affect for each modality, whereas many of the previous studies compare processing of positive and negative affective stimuli to neutral stimuli.

One methodological concern when examining responses to affective stimuli is the possibility that differences in physical characteristics of the stimuli can account for observed differences between the affective categories (e.g., Fecteau et al., 2007). Moreover, an exhaustive search of all

physical characteristics may not be feasible. Thus we cannot rule out an alternative explanation that the analyses are sensitive to stimulus attributes correlated to valence. However, tests indicated that for several features there was no significant difference. Strong evidence for measurement of valence would derive from successful cross-modal classification. The lack of cross-modal classification of valence implies that there was not consistent encoding of valence across modes captured by a sufficient number of voxels to enable above chance prediction. This could reflect a lack of general valence encoding in our paradigm or insufficient methods for finding such voxels if they were highly localized. The criticism that multivariate pattern analyses in this instance may be picking up on variables other than those targeted (valence in this case) holds for most studies and reflects the correlational nature of the data. However, we do know that our stimuli differed strongly on valence and several, but not all, concomitant variables were accounted for.

One notable study that did examine affective stimuli across modalities was reported by Peelen et al. (2010). Their study presented five emotions (anger, disgust, fear, happiness, and sadness) as portrayed in either videos of facial expressions, videos of body expressions, or auditory recordings of vocalized nonlinguistic interjections. Unlike our results, Peelen et al. (2010) found evidence for modality-general emotional processing using a searchlight method. The areas of common emotional processing were located in the medial prefrontal cortex and the left superior temporal sulcus. Thus, they found evidence of supramodal representations of these emotions in high-level brain areas linked to affective processing as well as attributions of mental states and theory-of-mind. There are several differences between our study and Peelen et al. (2010) that may account for these differences in results. First, we manipulated valence rather than specific emotions. According to the theory of core affect, valence and arousal are dimensions that contribute to emotion perception but do not in themselves constitute the emotional complex (Barrett and Bliss-Moreau, 2009). Thus, it may be that when affective responses are integrated into a more complex emotional percept, modality-general processing of affect comes into play. Second, our goal was to describe the internal representation of affect experienced in an incidental exposure paradigm and not localize supramodal representations underlying the identification of specific emotions. Supramodal representations may be more likely to occur when participants are intentionally processing the stimulus in order to identify the emotion and judge its intensity, activities in which they engaged in the Peelen et al. (2010) study. It may be that the act of formulating an appropriate affect-related response engages the visceromotor network and leads to modality-general affective processing. Finally, we have examined valence across modalities at the whole-brain level of analysis and thus cannot rule out a possibility that all three of our methods were not sensitive enough to a small number of modality-

general voxels.<sup>2</sup> These proposed explanations for the differences in our results and those of Peelen et al. (2010) are admittedly speculative and should be subject to testing by manipulation of task and stimulus elements of the design.

Previous studies of affective processing of sounds have often focused on responses in the auditory cortex. For example, auditory cortex has been examined for processing of affective sounds with near-infrared spectroscopy (Plichta et al., 2011). Speech content as well as speaker identity has been shown to be decoded from auditory cortex (Formisano et al., 2008). Moreover, vocal emotions (anger, sadness, neutral, relief, and joy) have been successfully decoded from bilateral voice-sensitive areas (Ethofer et al., 2009). Our study expands upon this literature by suggesting that valence can be successfully decoded from sound-sensitive areas as well.

Processing affective sounds has been previously shown to result in similar, but somewhat weaker physiological, self-report, and behavioral measures compared to processing affective pictures (Bradley and Lang, 2000; Mühl et al., 2011). Consistent with these previous findings, our behavioral results showed a greater difference in valence values for visual stimuli than for auditory stimuli. Our neuroimaging results also showed a trend for a stronger affective response to visual stimuli compared to affective response to auditory stimuli. It may be that the weaker results for sounds are due to the choice of specific exemplars, although the two sets were closely matched in normed ratings of valence (Table I). Alternatively, the pictures might have been more salient than sounds in our experiment due to presentation timing and perhaps the tendency of scanner noise to interfere with the auditory signals to some degree. Auditory stimuli durations were shortened from 5–6 s to 2 s, possibly affecting the naturalness of stimuli. However, the behavioral results showed a significant effect for valence for the pairs of shortened stimuli. Future research is needed to determine whether the auditory modality leads to reduced valence sensitivity.

Modality differences in the timing provide another possible explanation of why modality-specific valence processing was found in our study. It may be that voxels that appear to be differentially sensitive to valence for visual stimuli and auditory stimuli are actually differentially linked to the timing of valence processing. There is clear evidence that affective processing of visual stimuli is very fast and requires little exposure time (Codispoti et al., 2009; Junghofer et al., 2001, 2006). On the other hand, naturalistic auditory stimuli may require more time to identify and hence the affective processing of these stimuli may be delayed relative to visual stimuli. If this is the case, then different voxels tuned to auditory valence and

visual valence may reflect differences in the temporal processing of these stimuli rather than reflect strict modality differences. Investigations of the affective timing of stimuli presented in these two modalities should be helpful in resolving this issue.

In conclusion, we have shown the utility of examining the similarity structure of neural patterns in describing the internal representation of affective space. Modality-general and modality-specific valence hypotheses predicted distinctly different multidimensional patterning of our stimulus conditions. Our results clearly supported modality-specific representation for our particular experimental design. However, as we have discussed, changes in presentation and task variables might lead to modality-general valence representations, which could be detected by these analytic methods. Hence, we believe that similarity-based analyses of neuroimaging response patterns may provide a complementary and useful set of analytic tools for testing specific hypotheses concerning cognitive and affective representations.

## ACKNOWLEDGMENTS

We thank the anonymous reviewers for helpful comments.

## REFERENCES

- Abdi H, Dunlop JP, Williams LJ (2009): How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the Bootstrap and 3-way multidimensional scaling (DISTATIS). *Neuroimage* 45:89–95.
- Abdi H, Williams LJ, Valentin D, Bannani-Dosse M (2012): STATIS and DISTATIS: Optimum multitable principal component analysis and three way metric multidimensional scaling. *Wiley Interdiscip Rev Comput Stat* 4:124–167.
- Barrett LF, Bliss-Moreau E (2009): Affect as a psychological primitive. *Adv Exp Soc Psychol* 41, 167–218.
- Baucom LB, Wedell DH, Wang J, Blitzer DN, Shinkareva SV (2012): Decoding the neural representation of affective states. *Neuroimage* 59:718–727.
- Bishop CM (2006): *Pattern Recognition and Machine Learning*. New York: Springer.
- Bradley MM, Lang PJ (2000): Affective reactions to acoustic stimuli. *Psychophysiology* 37:204–215.
- Churchill NW, Oder A, Abdi H, Tam F, Lee W, Thomas C, Ween JE, Graham SJ, Strother S (2012): Optimizing preprocessing and analysis pipelines for single-subject fMRI. I. Standard temporal motion and physiological noise correction methods. *Hum Brain Mapp* 33:609–627.
- Codispoti M, Mazzetti M, Bradley MM (2009): Unmasking emotion: Exposure duration and emotional engagement. *Psychophysiology* 46:731–738.
- Connolly AC, Guntupalli JS, Gors J, Hanke M, Halchenko YO, Wu Y-C, Abdi H, Haxby JV (2012): The representation of biological classes in the human brain. *J Neurosci* 32:2608–2618.
- Dan-Glauser ES, Scherer KR (2011): The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behav Res Methods* 43: 468–477.

<sup>2</sup>Using the same searchlight analyses strategy as in Peelen et al. (2010) study on our data did not produce any results or indicate a trend for modality-general processing, however, our study is underpowered for searchlight analyses reported at a group level.

- Edelman S, Grill-Spector K, Kushnir T, Malach R (1998): Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology* 26:309–321.
- Ethofer T, Van De Ville D, Scherer K, Vuilleumier P (2009): Decoding of emotional information in voice-sensitive cortices. *Curr Biol* 19:1028–1033.
- Fecteau S, Belin P, Joanette Y, Armony JL (2007): Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage* 36:480–487.
- Formisano E, De Martino F, Bonte M, Goebel R (2008): “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science* 322:970–973.
- Jimura K, Poldrack RA (2012): Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia* 50:544–552.
- Junghofer M, Bradley MM, Elbert TR, Lang PJ (2001): Fleeting images: A new look at early emotion discrimination. *Psychophysiology* 38:175–178.
- Junghofer M, Sabatinelli D, Bradley MM, Schupp HT, Elbert TR, Lang PJ (2006): Fleeting images: Rapid affect discrimination in the visual cortex. *Neuroreport* 17:225–229.
- Kassam KS, Markey AR, Cherkassky VL, Loewenstein G, Just MA (2013): Identifying emotions on the basis of neural activation. *PLoS One* 8:e66032.
- Kherif F, Poline JB, Meriaux S, Benali H, Flandin G, Brett M (2003): Group analysis in functional neuroimaging: Selecting subjects using similarity measures. *Neuroimage* 20:2197–2208.
- Kotz SA, Kalberlah C, Bahlmann J, Friederici AD, Haynes JD (2013): Predicting vocal emotion expressions from the human brain. *Hum Brain Mapp* 34:1971–1981.
- Kriegeskorte N, Mur M, Bandettini P (2008a): Representational similarity analysis—Connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:1–28.
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008b): Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60:1126–1141.
- Lakens D, Fockenberg DA, Lemmens KPH, Ham J, Midden CJH (2013): Brightness differences influence the evaluation of affective pictures. *Cogn Emot* 27:1225–1246.
- Lang PJ, Bradley MM, Cuthbert BN (2008): International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. Gainesville, FL: University of Florida.
- Lang PJ, Bradley MM, Fitzsimmons JR, Cuthbert BN, Scott JD, Moulder B, Nangia V (1998): Emotional arousal and activation of the visual cortex: An fMRI analysis. *Psychophysiology* 35: 199–210.
- Lavit C, Escoufier Y, Sabatier R, Traissac P (1994): The ACT (STATIS method). *Comput Stat Data Anal* 18:97–119.
- Misaki M, Kim Y, Bandettini PA, Kriegeskorte N (2010): Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* 53:103–118.
- Mourao-Miranda J, Volchan E, Moll J, de Oliveira-Souza R, Oliveira L, Bramati I, Gattass R, Pessoa L (2003): Contributions of stimulus valence and arousal to visual activation during emotional perception. *Neuroimage* 20:1955–1963.
- Mühl C, van den Broek EL, Brouwer A-M, Nijboer F, van Wouwe N, Heylen D (2011): Affective brain-computer interfaces. In: D’Mello S, Graesser A, Schuller B, Martin J-C, editors. *Affective Computing and Intelligent Interaction*. Memphis, TN: Springer. pp 235–245.
- O’Toole AJ, Jiang F, Abdi H, Penard N, Dunlop JP, Parent MA (2007): Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *J Cognitive Neurosci* 19:1735–1752.
- Peelen MV, Atkinson AP, Vuilleumier P (2010): Supramodal representations of perceived emotions in the human brain. *J Neurosci* 30:10127–10134.
- Pereira F, Mitchell T, Botvinick M (2009): Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage* 45:S199–S209.
- Pessoa L, Padmala S (2007): Decoding near-threshold perception of fear from distributed single-trial brain activation. *Cereb Cortex* 17:691–701.
- Plichta MM, Gerdes ABM, Alpers GW, Harnisch W, Brill S, Wieser MJ, Fallgatter AJ (2011): Auditory cortex activation is modulated by emotion: A functional near-infrared spectroscopy (fNIRS) study. *Neuroimage* 55:1200–1207.
- Robert P, Escoufier Y (1976): A unifying tool for linear multivariate statistical methods: The RV-coefficient. *Appl Stat* 25:257–265.
- Russell J (2003): Core affect and the psychological construction of emotion. *Psychol Rev* 110:145–172.
- Russell J, Barrett LF (1999): Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *J Pers Soc Psychol* 76:805–819.
- Said CP, Moore CD, Engell AD, Todorov A, Haxby JV (2010): Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J Vis* 10:1–12.
- Shinkareva SV, Malave VL, Just MA, Mitchell TM (2012): Exploring commonalities across participants in the neural representation of objects. *Hum Brain Mapp* 33:1375–1383.
- Shinkareva SV, Malave VL, Mason RA, Mitchell TM, Just MA (2011): Commonality of neural representations of words and pictures. *Neuroimage* 54:2418–2425.
- Shinkareva SV, Mason RA, Malave VL, Wang W, Mitchell TM, Just MA (2008): Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS One* 3:e1394.
- Shinkareva SV, Wang J, Wedell DH (2013): Examining similarity structure: Multidimensional scaling and related approaches in neuroimaging. *Comput Math Methods Med* 2013:796183.
- Sitaram R, Lee S, Ruiz S, Rana M, Veit R, Birbaumer N (2011): Real-time support vector classification and feedback of multiple emotional brain states. *Neuroimage* 56:753–765.
- Viinikainen M, Kätsyri J, Sams M (2012): Representation of perceived sound valence in the human brain. *Hum Brain Mapp* 33:2295–2305.
- Wang J, Baucom L, Shinkareva SV (2013): Decoding abstract and concrete concept representations based on single-trial fMRI data. *Hum Brain Mapp* 34:1133–1147.
- Yuen KSL, Johnston SJ, De Martino F, Sorger B, Formisano E, Linden DEJ, Goebel R (2012): Pattern classification predicts individuals’ responses to affective stimuli. *Transl Neurosci* 3: 278–287.