Video Article

Cross-Modal Multivariate Pattern Analysis

Kaspar Meyer[1], Jonas T. Kaplan[1]

[1]Brain and Creativity Institute and Department of Psychology, University of Southern California

Correspondence to: Kaspar Meyer at kaspar.meyer@usc.edu

URL: https://www.jove.com/video/3307
DOI: doi:10.3791/3307

Keywords: Neuroscience, Issue 57, perception, sensory, cross-modal, top-down, mental imagery, fMRI, MRI, neuroimaging, multivariate pattern analysis, MVPA

Date Published: 11/9/2011

Citation: Meyer, K., Kaplan, J.T. Cross-Modal Multivariate Pattern Analysis. *J. Vis. Exp.* (57), e3307, doi:10.3791/3307 (2011).

Abstract

Multivariate pattern analysis (MVPA) is an increasingly popular method of analyzing functional magnetic resonance imaging (fMRI) data[1-4]. Typically, the method is used to identify a subject's perceptual experience from neural activity in certain regions of the brain. For instance, it has been employed to predict the orientation of visual gratings a subject perceives from activity in early visual cortices[5] or, analogously, the content of speech from activity in early auditory cortices[6].

Here, we present an extension of the classical MVPA paradigm, according to which perceptual stimuli are not predicted within, but across sensory systems. Specifically, the method we describe addresses the question of whether stimuli that evoke memory associations in modalities other than the one through which they are presented induce content-specific activity patterns in the sensory cortices of those other modalities. For instance, seeing a muted video clip of a glass vase shattering on the ground automatically triggers in most observers an auditory image of the associated sound; is the experience of this image in the "mind's ear" correlated with a specific neural activity pattern in early auditory cortices? Furthermore, is this activity pattern distinct from the pattern that could be observed if the subject were, instead, watching a video clip of a howling dog?

In two previous studies[7,8], we were able to predict sound- and touch-implying video clips based on neural activity in early auditory and somatosensory cortices, respectively. Our results are in line with a neuroarchitectural framework proposed by Damasio[9,10], according to which the experience of mental images that are based on memories - such as hearing the shattering sound of a vase in the "mind's ear" upon seeing the corresponding video clip - is supported by the re-construction of content-specific neural activity patterns in early sensory cortices.

Video Link

The video component of this article can be found at https://www.jove.com/video/3307/

Protocol

1. Introduction

Multivariate pattern analysis (MVPA) is an increasingly popular method of analyzing functional magnetic resonance imaging (fMRI) data[1-4]. Typically, the method is used to identify a subject's perceptual experience from neural activity in certain regions of the brain. For instance, it has been employed to predict the orientation of visual gratings a subject perceives from activity in early visual cortices[5] or, analogously, the content of speech from activity in early auditory cortices[6]. In this video article, we describe a novel application of MVPA which adds an extra twist to this basic, intra-modal paradigm. In this approach, perceptual stimuli are predicted not within, but across sensory systems.

2. Multivariate Pattern Analysis

Although the MVPA method by now is well established within the neuroimaging realm, we will start by pointing out the key differences between MVPA and conventional, univariate fMRI analysis. To this end, consider the following example of how the two methods go about examining neural activity in the visual cortex during a simple visual task (Video Clip 1):

1. A subject is presented with two different visual stimuli, for instance, an image of an orange and an image of an apple.
2. Both stimuli induce a specific pattern of neural activity in the primary visual cortex, symbolized here by the activation levels of six hypothetic voxels. (Of course, activity patterns induced by a single presentation of the orange or apple images in reality would be very noisy; consider the illustrated patterns as averages resulting from a large number of trials.)
3. In conventional fMRI analysis, there are essentially two ways in which these patterns can be analyzed. First, one can focus on the average level of activity across the entire region of interest.
4. In the example given, the difference in average activity levels is not significant, so that the patterns corresponding to the two stimuli cannot be distinguished from this point of view.

Copyright © 2011 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

5. Another way to analyze the two patterns is to establish a subtraction contrast between them: for each voxel, the activation level during the "apple" condition is subtracted from the activation level during the "orange" condition. The resulting difference can then be visualized for each voxel on a whole-brain contrast image.
6. Again, however, these differences may be small and may reach the required statistical criterion only for very few voxels.
7. This is where the decisive advantage of MVPA comes into play: its superior power derives from the fact that, unlike univariate analysis methods, it considers the activation levels of all voxels simultaneously and thus is able to detect patterns within them. While, as mentioned, only few of the activation differences may be significant when considered in isolation, the two patterns, when considered in their entirety, may indeed by statistically different.

There is a second major difference between conventional fMRI analysis and MVPA (Video Clip 2). The former method typically attempts to demonstrate a statistical dependency between certain sensory stimuli and certain brain activity patterns in a "forward manner"; in other words, it asks question of the type: "Will two different visual stimuli, e.g. the picture of a face and the picture of a house, lead to different activity levels in a specific region of interest, e.g. the fusiform face area?" By contrast, the success of MVPA is typically expressed in terms of "reverse inference" or "decoding"; the typical question is of the type: "Based on the pattern of neural activity in a specific region of interest (e.g. the primary visual cortex), can one predict whether a subject perceives stimulus A, e.g. an orange, or stimulus B, e.g. an apple?" Note, however, that the direction in which the correlation between the perceptual stimuli and brain activity is mapped does not matter from a statistical point of view: it is equivalent to say that two stimuli lead to distinct activity patterns in a given brain region and to say that the activity pattern in that brain region permits prediction of the inducing stimulus[11]. In other words, the sensitivity of MVPA is superior to that of univariate analyses because it considers several voxels simultaneously, and not because it proceeds in an inverse direction.

The following steps illustrate how a typical MVPA paradigm would address the question of whether seeing an apple induces a different pattern of neural activity in primary visual cortex than seeing an orange (Video Clip 3):

1. fMRI data are acquired while a subject sees large number of apple and orange stimuli.
2. The acquired data are divided into a training data set and a testing data set. Unlike in the present example, the training data set is often chosen to be larger than the testing data set, as the performance of the classifier can be expected to increase with the number of training trials.
3. The data from the training set are entered into a pattern classifier. Using one of several possible mathematical algorithms, the classifier attempts to detect features in the neural patterns that distinguish the two stimulus types from one another. A type of classifier commonly used (also in our own previous studies) are so-called support vector machines; for further detail, the reader is referred to the reviews mentioned in the introduction.
4. After the classifier has been trained on the training trials, it is provided with the testing data. The individual trials from the testing data set are unlabeled; in other words, the classifier does not "know" whether a pattern comes from an "apple" or an "orange" trial.
5. Based on the consistencies it was able to detect in the training data set, the classifier attributes the most likely label to each of the testing trials.
6. For each pattern, the classifier "guess" can be compared with the correct stimulus label.
7. If the classifier was unable to detect any consistent differences between the patterns induced by the two stimuli, its performance should be at chance level; for the two-way discrimination given in the example, this would correspond to 50% correct labels. A prediction performance significantly above this value indicates that there are indeed consistent differences between the two types of stimuli.

Note that it is crucial that the training and testing data sets be independent from one another. Only if this is the case can any conclusions be drawn as to the generalizability of the patterns derived from the training set. MVPA studies often assess classifier performance using a cross-validation paradigm (Video Clip 4). Assume that an MVPA experiment consists of eight functional runs. In the first cross-validation step, a classifier is trained on the data from runs 1 through 7 and tested on the data from run 8. In the second step, the classifier is then trained on runs 1 through 6 as well as run 8, and subsequently tested on run 7. Following this schema, eight cross-validation steps are carried out, with each run serving as test run exactly once. Overall classifier performance is calculated as the average of the performances on the individual cross-validation steps. While this procedure guarantees independent training and testing data sets on each step, it also maximizes the overall number of testing trials, which can be of advantage when assessing the statistical significance of the classifier's performance.

There are freely available software packages on the internet to perform MVPA; two examples are PyMVPA[12] (based on Python; http://www.pymvpa.org) and the toolbox offered by the Princeton Neuroscience Institute (based on Matlab; http://code.google.com/p/princeton-mvpa-toolbox/).

3. Cross-Modal MVPA and the Framework of Convergence-Divergence Zones

As mentioned in the introduction, experimental paradigms like the one just described have been used successfully to predict perceptual stimuli from neural activity in corresponding sensory cortices, in other words, visual stimuli based on activity in visual cortices and auditory stimuli based on activity in auditory cortices. Here, we present an extension of this basic concept. Specifically, we hypothesized that it should be possible to predict perceptual stimuli not only within, but across modalities. Sensory perception is intricately linked to the recall of memories; for example, a visual stimulus that has a strong auditory implication, such as the sight of a glass vase shattering on the ground, will automatically trigger in our "mind's ear" images that share similarities with the auditory images we experienced on previous encounters with breaking glass. According to a framework introduced by Damasio more than two decades ago[9,10], the memory association between the sight of the vase and the corresponding sound images is stored in so-called convergence-divergence zones (CDZs; Video Clip 5). CDZs are neuron ensembles in the association cortices which receive converging bottom-up projections from various early cortical areas (via several hierarchical levels) and which, in turn, send back divergent top-down projections to the same cortical sites. Due to the convergent bottom-up projections, CDZs can be activated by perceptual representations in multiple modalities - for instance, both by the sight and the sound of a shattering vase; due to the divergent top-down projections, they can then promote the reconstruction of associated images by signaling back to the early cortices of additional modalities. Damasio emphasized the latter point: activating CDZs in association cortices would not be sufficient for the conscious recall of an image from memory; only once CDZs would reconstruct explicit neural representations in early sensory cortices would the image be

Copyright © 2011 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

consciously experienced. Thus, the framework predicts a specific sequence of neural processing in response to a (purely) visual stimulus that implies sound (Video Clip 6):

1. The stimulus first induces a specific pattern of neural activity (red rectangles) in the early visual cortices.
2. Via convergent bottom-up projections, the neurons in early visual cortices project forward to a first level of CDZs (CDZ_1s). The convergent pattern of connectivity permits the CDZ_1s to detect certain patterns of activity in the early visual cortices. Depending on the exact pattern, a CDZ may or may not become activated. CDZs thus act as feature extractors. In this example, two CDZ_1s become activated (as indicated by the red color), while the third is not triggered by the specific activity pattern in the corresponding sector of the early visual cortices.
3. CDZ_1s send convergent bottom-up projections to CDZ_2s; therefore, just as CDZ_1s detected certain patterns of activity in the early visual cortices, CDZ_2s are able to detect patterns of activity among CDZ_1s. Several CDZ_2s may become activated by the specific configuration of activated CDZ_1s; for reasons of simplicity, only a single CDZ_2 is depicted here. In the example, the activity pattern among CDZ_1s is adequate to activate this CDZ_2.
4. Notably, CDZ_1s do not only project forward to CDZ_2s but also back to the early cortices (blue arrows). These top-down signals may complete the (possibly noisy) activity pattern initially induced by the stimulus (blue rectangle). Generally, red color denotes bottom-up activations, while blue color represents top-down activations.
5. Via several additional levels of CDZs, the CDZ_2s project forward to CDZ_ns in higher-order association cortices (dashed arrow). One or several CDZ_ns may respond to the specific visual stimulus in question (only a single one is depicted).
6. Again, it should be noted that the CDZ_2s also signal backward to CDZ_1s which, in turn, may further modify the pattern originally induced in the early visual cortices.
7. The CDZ_ns signal back to the CDZ_2s of all modalities. In the visual cortices, this may lead to the completion of activity patterns in lower-level CDZs. In the auditory cortices, a neural pattern will be constructed - first at the level of CDZ_2s and CDZ_1s, ultimately in the early auditory cortices - which permits the conscious experience of an auditory image associated with the visually presented stimulus. Note that there is also top-down signaling to the somatosensory modality, although to lesser extent than for the auditory modality. This reflects the fact that almost any visual stimulus has some tactile association to it. As the visual stimulus in the current example is assumed to specifically imply sound, however, top-down signaling to the auditory cortices is more extensive.

Based on the proposed sequence of neural processing, the framework makes a specific prediction: visual stimuli containing objects and events that strongly imply sound should evoke neural activity in early auditory cortices. Moreover, the auditory activity patterns should be stimulus-specific; in other words, a video clip of a shattering vase should induce a different pattern than a clip of a howling dog. If this prediction were correct, then we should indeed be able to perform MVPA cross-modally: for instance, we should be able to predict, based exclusively on the neural activity fingerprint in early auditory cortices, whether a person is seeing a shattering vase or a howling dog (Video Clip 7). Naturally, analogous paradigms invoking information transfer among other sensory modalities should also be successful. For instance, if the video clips shown to a subject implied touch rather than sound, we should be able to predict those clips from the activity patterns they elicit in early somatosensory cortex.

4. Stimuli

The general paradigm of an MVPA study was described in Section 2. Our approach is different from previous studies in that it attempts to perform MVPA across sensory systems and therefore uses stimuli that are specifically designed to have implications in a sensory modality other than the one in which they are presented. In one previous study, for instance, we recorded neural activity from primary somatosensory cortex while subjects watched 5-second video clips of everyday objects being manipulated by human hands[8] (Video Clip 8 and Video Clip 9). In another study, we investigated neural activity in early auditory cortices while subjects viewed video clips that depicted objects and events that strongly implied sound[7] (Video Clip 10 and Video Clip 11). However, according to the CDZ framework, sensory stimuli of all modalities can potentially be employed in this general paradigm, as long as they have implications in additional modalities.

5. Regions of Interest

Generally, the regions of interest for a neuroimaging study can be determined either functionally or anatomically. We believe that in the experimental paradigm we describe here, anatomical localizers are more suitable for two reasons. First, it is not trivial to functionally define the primary or early cortices of a given sensory modality (with the possible exception of the primary visual cortex), as processing of perceptual stimuli presented to the subject in that modality typically will not be limited to these areas. For instance, it would be difficult to define the primary somatosensory cortex by applying touch to a subject's hands, as the activity induced by this procedure would, in all likelihood, spread to somatosensory association cortices as well. Second, a functional localizer may not label all the voxels that could potentially contribute to the classifier's performance: it has been shown that areas that do not show net activation in response to sensory stimuli in the classical sense (i.e., regions that do not appear on a contrast image [stimulation vs. rest]) can contain information about the stimuli nonetheless[13,14]. For these two reasons, we advocate the use of anatomically defined regions of interest whenever macroscopic landmarks allow for this; for example, the gross anatomy of the postcentral gyrus represents a reasonable approximation of the primary somatosensory cortex, and we used this to define the region of interest in our somatosensory study[8] (Figure 1).

6. Subjects

The subject samples in MVPA studies tend to be smaller than in conventional fMRI studies, as the analysis can be performed at the single-subject level. Of course, this does not prevent the experimenter from subsequently analyzing the results of the individual subjects at the group level as well. In the two studies mentioned earlier, for example, we conducted t-tests on the individual subject results in order to assess their significance at the group level. Each study involved eight subjects; although this must be considered a very small subject sample for parametric testing, we did find many of the discriminations we assessed to be significant (see below).

Copyright © 2011 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

7. Representative Results:

As mentioned, in two previous studies we aimed to predict sound-implying video clips based on neural activity in early auditory cortices[7] (see Figure 2 for the mask used in this study) and touch-implying video clips based on activity in primary somatosensory cortices[8]. This attempt was successful: in both studies, an MVPA classifier performed above the chance level of 50% for all possible two-way discriminations between stimulus pairs (n = 36 in the auditory study, given there were 9 different stimuli; n = 10 for the somatosensory study, given there were 5 different stimuli). In the auditory study, 26 out of the 36 discriminations reached statistical significance; in the somatosensory study, this was the case for 8 out of the 10 discriminations (two-tailed t-tests, n = 8 in both studies; Figure 3).

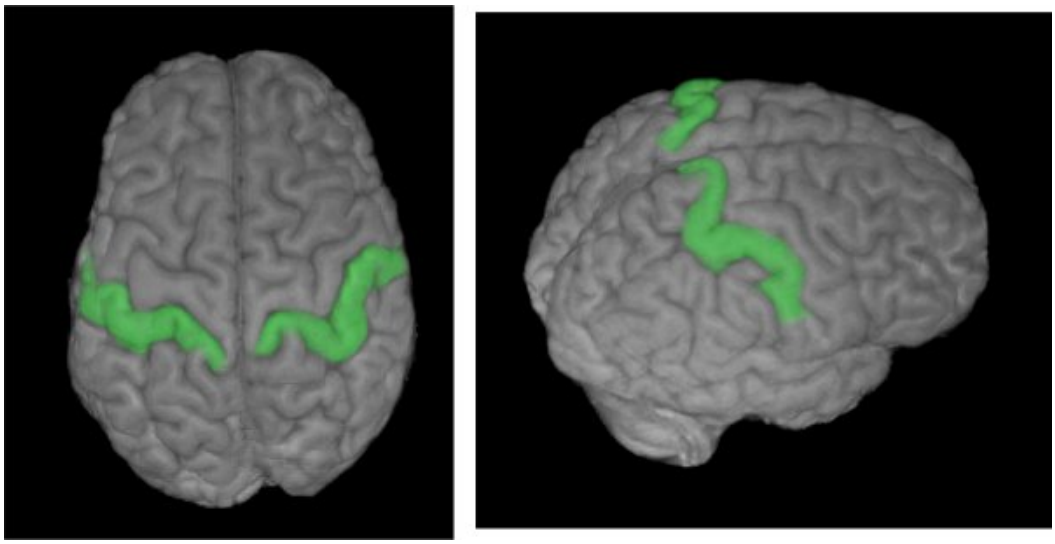

Figure 1. Extent of the anatomically defined mask of primary somatosensory cortex, as used in Meyer *et al.*, 2011. A classifier algorithm was able to predict touch-implying video clips from brain activity patterns restricted to the demarcated region. Reproduced with permission from Oxford University Press.

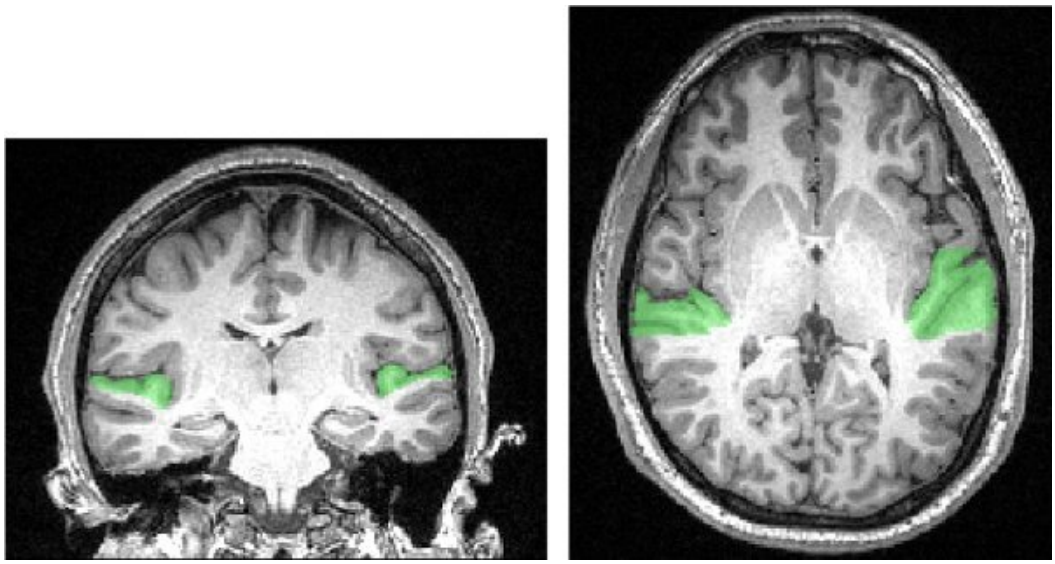

Figure 2. Extent of the anatomically defined mask of early auditory cortices, as used in Meyer *et al.*, 2010. A classifier algorithm was able to predict (silent) sound-implying video clips from brain activity patterns restricted to the demarcated region. Reproduced with permission from Nature Publishing Group.

Copyright © 2011 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

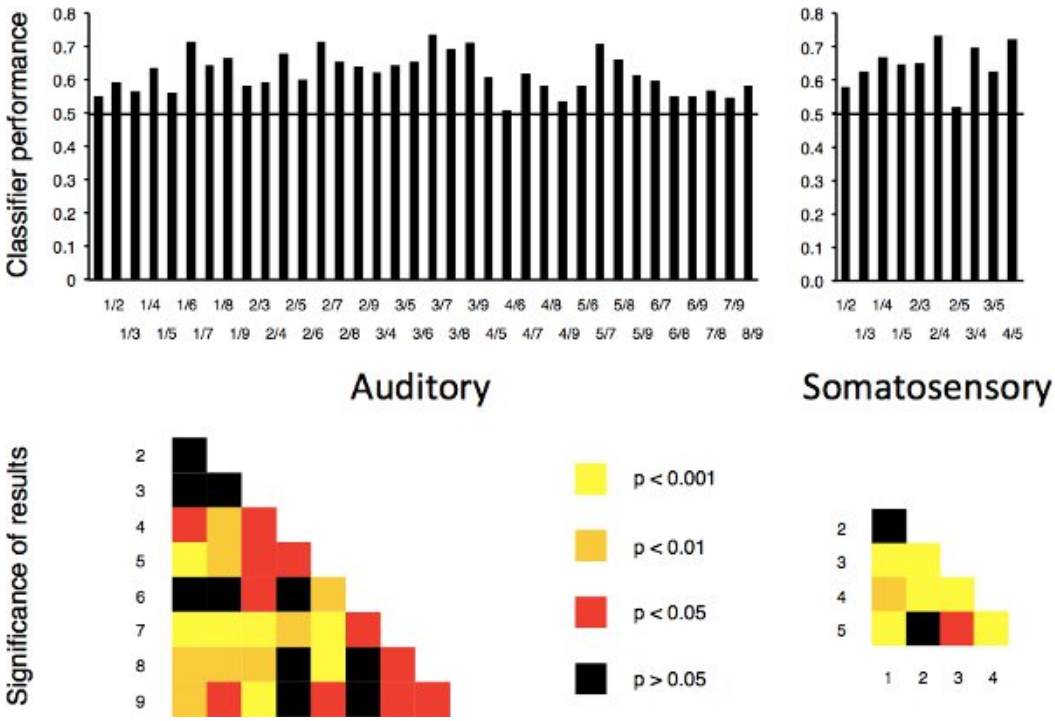

Figure 3. Summary of the results of our previous cross-modal MVPA studies. A classifier was used to predict visual stimuli that implied either sound or touch from activity in the early auditory or primary somatosensory cortices, respectively. Top panels: in both studies, prediction performance was above the chance level of 0.5 for all two-way discriminations between pairs of stimuli. Bottom panels: in the auditory study, classifier performance reached statistical significance for 26 of the 36 discriminations; in the somatosensory study, this was the case for 8 of the 10 discriminations. Reproduced with permission from Nature Publishing Group and Oxford University Press.

Discussion

The findings of our previous studies demonstrate that cross-modal MVPA is a useful tool to study the neural correlates of mental images experienced in the "mind's ear" and the "mind's touch". Specifically, the results show that the content of such images is correlated with neural activity in the early auditory and somatosensory cortices, respectively, providing direct empirical support for Damasio's framework of convergence-divergence zones.

The basic paradigm we describe can be extended in a number of ways. Most obviously, similar studies can be carried out using different combinations of sensory modalities. In this respect, it is possible that establishing certain cross-modal associations prior to the experiment may increase the chances of success. For instance, to study cross-modal representations in olfactory cortices, one could prime subjects by exposing them simultaneously to the sight and the smell of a number of food items. Shortly after, inside the fMRI scanner, the olfactory memories of the smells would be triggered by visual cues, and MVPA could be used in an attempt to assign the trials to the correct food items based solely on activity in olfactory cortices.

Another question of interest is whether the cross-modally induced patterns of activity in the early sensory cortices bear similarities with the patterns that are induced when sound or touch is actually experienced; in other words, does seeing the glass vase shatter invoke a similar pattern of neural activity in early auditory cortices as actually hearing the same event? This question, again, can be addressed through MVPA: is a classifier that has been trained on data recorded while subjects heard certain sounds able to correctly discriminate data recorded while the subjects watched corresponding video clips? In our auditory study, we did attempt such a classification, but the results were marginal (see Fig. 3 in ref. 7). In that study, however, the auditory associations of the participants were not controlled in any way; in other words, we do not know how similar the auditory images each subject associated with videos were to the audio tracks used to train the classifier. Again, it might be interesting to address the same question after specific cross-modal associations have been primed in the subjects, as described above for a visuo-olfactory association. This would permit to control more reliably the mental experience of the subjects during the video trials and thus might increase prediction performance of the classifier.

To conclude, we have introduced an extension to the classical MVPA paradigm by showing that stimuli can be predicted not only within, but also across sensory modalities. Thus, we show that the use of MVPA is not limited to investigating the correlates of perceptual representations induced directly by external sensory stimulation. Rather, MVPA can also assess the neural substrate of mental images that are triggered internally: in accordance with Damasio's convergence-divergence zone framework, our findings suggest that the conscious experience of mental images that are reconstructed based on memories is correlated with content-specific neural representations in the early sensory cortices.

Disclosures

No conflicts of interest declared.

Copyright © 2011 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License

Acknowledgements

This work was supported by grants by the Mathers Foundation and the National Institutes of Health (grant number 5P50NS019632-27) to Antonio and Hanna Damasio.

References

1. Haynes, J.-D. & Rees, G. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* **7**, 523-534 (2006).
2. Norman, K.A., Polyn, S.M., Detre, G.J., & Haxby, J.V. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends. Cogn. Sci.* **10**, 424-430 (2006).
3. Kriegeskorte, N. & Bandettini, P. Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage.* **38**, 649-662 (2007).
4. Mur, M., Bandettini, P.A., & Kriegeskorte, N. Revealing representational content with pattern-information fMRI - an introductory guide. *Soc. Cogn. Affect. Neurosci.* **4**, 101-109 (2009).
5. Kamitani, Y. & Tong, F. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* **8**, 679-685 (2005).
6. Formisano, E., De Martino, F., Bonte, M. & Goebel, R. "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science.* **322**, 970-973 (2008).
7. Meyer, K., Kaplan, J.T., Essex, R., Webber, C., Damasio, H., & Damasio, A. Predicting visual stimuli on the basis of activity in auditory cortices. *Nat. Neurosci.* **13**, 667-668 (2010).
8. Meyer, K., Kaplan, J.T., Essex, R., Damasio, H., & Damasio, A. Seeing touch is correlated with content-specific activity in primary somatosensory cortex. *Cereb. Cortex.* **21**, 2113-2121 (2011).
9. Damasio, A.R. Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition.* **33**, 25-62 (1989).
10. Meyer, K. & Damasio, A. Convergence and divergence in a neural architecture for recognition and memory. *Trends. Neurosci.* **32**, 376-382 (2009).
11. Friston, K.J. Modalities, modes, and models in functional neuroimaging. *Science.* **326**, 399-403 (2009).
12. Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J.V., & Pollmann, S. PyMVPA: a Python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics.* **7**, 37-53 (2009).
13. Harrison, S.A. & Tong, F. Decoding reveals the contents of visual working memory in early visual areas. *Nature.* **458**, 632-635 (2009).
14. Serences, J.T., Ester, E.F., Vogel, E.K., & Awh, E. Stimulus-specific delay activity in human primary visual cortex. *Psych Sci.* **20**, 207-214 (2009).

Copyright © 2011 Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License