Stereoscopic depth increases intersubject correlations of brain networks

Michael Gaebler a,b,c,1, Felix Biessmann d,e,f,⁎,1, Jan-Peter Lamke a, Klaus-Robert Müller d,e,g, Henrik Walter a, Stefan Hetzer a,g,h

1 Charité – Universitätsmedizin Berlin, Germany
2 Max Planck Institute for Human Cognitive & Brain Sciences, Leipzig, Germany
3 Universität Leipzig, Germany
4 Korea University, Seoul, Republic of Korea
5 Technische Universität Berlin, Germany
6 Amazon Development Center, Berlin, Germany
7 Bernstein Center for Computational Neuroscience, Berlin, Germany
8 Berlin Center for Advanced Neuroimaging, Berlin, Germany
9 Available online xxxx
10 Article info

12 A R T I C L E  I N F O

13 Article history:
14 Accepted 4 June 2014
15 Available online xxxx
16 Keywords:
17 Intersubject correlations
18 Canonical correlation analysis (CCA)
19 Natural viewing
20 3D movies
21 fMRI
22 Immersion

19 A B S T R A C T

Three-dimensional movies presented via stereoscopic displays have become more popular in recent years aiming at a more engaging viewing experience. However, neurocognitive processes associated with the perception of stereoscopic depth in complex and dynamic visual stimuli remain understudied. Here, we investigate the influence of stereoscopic depth on both neurophysiology and subjective experiences. Using multivariate statistical learning methods, we compare the brain activity of subjects when freely watching the same movies in 2D and in 3D. Subjective reports indicate that 3D movies are more strongly experienced than 2D movies. On the neural level, we observe significantly higher intersubject correlations of cortical networks when subjects are watching 3D movies relative to the same movies in 2D. We demonstrate that increases in intersubject correlations of brain networks can serve as neurophysiological marker for stereoscopic depth and for the strength of the viewing experience.

© 2014 The Authors. Published by Elsevier Inc.

1. Introduction

Because of the horizontal separation of the eyes in the head, each eye receives slightly different images of the world. The cerebral cortex integrates these subtle differences, so-called binocular disparities, with additional perceptual cues and cognitive factors. It thus enables us to perceive stereoscopic depth and to experience three-dimensional space (Parker, 2007). For almost two centuries, this perceptual phenomenon has been used to induce an illusion of depth by presenting two offset images separately to each eye (Crone, 1992; Wheatstone, 1838). In recent years, the presentation of moving images increasingly employs stereoscopic depth and 3D stimuli have entered not only television, computer games, and cinema but also professional environments (Parker, 2007). For almost two centuries, this perceptual phenomenon has been used to induce an illusion of depth by presenting two offset images separately to each eye (Parker, 2007).

...
2. Material and methods

2.1. Participants

Twenty-six healthy participants were recruited through ads in public spaces and through university mailing lists. Participants did not report past or current neurological disease, psychiatric disorder, or drug abuse, did understand German, and did not fulfill standard MRI exclusion criteria. Prior to scanning, participants were thoroughly informed and gave their written consent. The experiment was approved by the local ethics committee. One participant was excluded due to intense sleepiness. Participants that entered the analysis were 25 young adults (12 males, 13 females) at an average age of 26.7 years (SD: 3.5, range 21–35). They were all right-handed according to self-report (Oldfield, 1971) (mean: 90, SD: 14) and had been on average 3.0 times (SD: 2.3, range 0–10) to a 3D cinema, as assessed after participation in the experiment. All subjects had normal vision: 7 corrected by eyeglasses, 2 by contact lenses, and 16 did not need correction. For those who needed correction, mean dioptric values were −5.8 to +2.5). They were all right-handed according to self-report (Oldfield, 1971) (mean: 90, SD: 14) and had been on average 3.0 times (SD: 2.3, range 0–10) to a 3D cinema, as assessed after participation in the experiment. All subjects had normal vision: 7 corrected by eyeglasses, 2 by contact lenses, and 16 did not need correction. For those who needed correction, mean dioptric values were −5.8 to +2.5).

2.2. Stimuli

Stimuli were 14 different videos of 42.5 s length each: content length was 40.5 s, preceded by 2 s of black screen without fixation cross for visual adjustment and to avoid distortions induced by codec and presentation software. Videos were presented at 30 frames per second, resulting in a total number for each stimulus of 1275 frames at size 768 × 576 pixels (~28.8 × 21.6°) on each eye. The videos, which were all originally shot using stereoscopic recording equipment, were acquired over the internet. Their content varied from, for example, a calm time lapse montage of a blossoming flower (http://www.stereomaker.net/sample/index.html, accessed March 20, 2013) to a rapid car rally, filmed by onboard cameras (http://alesco.cz/, accessed December 8, 2012; see Table 1). Videos were edited using VirtualDub 1.9.11 (http://www.virtualdub.org/) and encoded using the XVID codec. Every movie was shown twice: in the 3D condition, stereoscopic depth was induced by presenting the two binocular perspectives of the scene to the corresponding eyes, while in the 2D condition, the same stimulus (left eye) was delivered to both eyes. As stimulus order was pseudo-randomized for each participant, novelty effects were balanced and stimulus characteristics were controlled for. The latter involve not only statistical properties like brightness, contrast, color, or motion but also more subjective stimulus features like personal preference for the movie content. Individual videos were interspersed with 20 s blocks of fixation and the first video presentation was preceded by a 30 s baseline fixation block. Presentation software version 14.9 (Neurobehavioral Systems, Inc., Albany, CA), a stereo adapter, and MR-compatible video goggles with a native resolution of 800 × 600 pixels and a color depth of 32 bit (VisualSystem, NordicNeuroLab, Bergen, Norway) were used for stimulus presentation. Careful adjustment of the goggle system and its built-in dioptric correction prior to scanning ensured optimal stimulus visibility.

2.3. Procedure

All participants were naive with respect to content and category of the stimuli. Before entering the scanner, they completed a short demographic questionnaire. They were instructed to attentively watch the presented movies and answer four questions after each presentation. On a Likert-type scale from 1 to 7 with labeled extremes they rated (a) valence (1: negative, 7: positive), (b) arousal (1: weak, 7: strong), (c) immersion into the presented movie (1: weak, 7: strong), and (d) awareness of the MRI environment (1: weak, 7: strong) for each stimulus. Ratings were collected using an MRI-compatible button box with three buttons (Current Designs, Inc., Philadelphia, PA, USA). Using the index, middle, and ring fingers of the right hand, participants could shift the colored circle, which indicated their current selection, before confirming their choice using the middle button. They were not restricted in their time to answer each question. After the scanning session, participants answered questions regarding their experience with and opinion about 3D movies, before they completed a questionnaire of immersive tendencies (QIT; Cronbach’s α = .80), which assessed their personal habits of being drawn into apparent realities like novels or movies (Scheuchenpflug et al., 2003).

Table 1

<table>
<thead>
<tr>
<th>No.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ride through a city in an oldtimer car</td>
</tr>
<tr>
<td>2</td>
<td>Time lapse movie of a pink flower opening and closing its bloom</td>
</tr>
<tr>
<td>3</td>
<td>Flock of dolphins swimming through underwater plants</td>
</tr>
<tr>
<td>4</td>
<td>Police sheriff and woman exploring a dark alley</td>
</tr>
<tr>
<td>5</td>
<td>Skateboarders doing tricks in a skateboard hall</td>
</tr>
<tr>
<td>6</td>
<td>Mountain bikers jumping over gaps in a dirt course</td>
</tr>
<tr>
<td>7</td>
<td>Three people fishing and exchanging money, two leaving in a canoe</td>
</tr>
<tr>
<td>8</td>
<td>Race car rally through the woods</td>
</tr>
<tr>
<td>9</td>
<td>Roller coaster ride</td>
</tr>
<tr>
<td>10</td>
<td>Skydive with the jump from the plane, free fall, and landing</td>
</tr>
<tr>
<td>11</td>
<td>Surfer standing on his board and riding a wave</td>
</tr>
<tr>
<td>12</td>
<td>Individual manatee, then a flock of manatees under water</td>
</tr>
<tr>
<td>13</td>
<td>People jumping over a cliff in wingsuit costumes</td>
</tr>
<tr>
<td>14</td>
<td>Scenes from a Graffiti and BMX event</td>
</tr>
</tbody>
</table>

Please cite this article as: Gaebler, M., et al., Stereoscopic depth increases intersubject correlations of brain networks, NeuroImage (2014), http://dx.doi.org/10.1016/j.neuroimage.2014.06.008
2.4. Data acquisition

MR imaging was performed on a Siemens TIM Trio 3T MR scanner with a standard 12-channel head coil (Siemens Medical Solutions, Erlangen, Germany) at the Berlin Center for Advanced Neuroimaging. A T1-weighted image was acquired as a high-resolution anatomical reference using a 3D-MPRAGE sequence with isotropic voxels of 1 mm³. 192 sagittal slices, repetition time (TR) 1900 ms, echo time (TE) 2.52 ms, flip angle 9°, field of view (FoV) 25 × 25 cm². T2-weighted gradient-echo echo-planar images (EPI) were collected for whole-brain functional imaging with isotropic voxels of 2.5 × 2.5 × 2.5 mm³ using a standard 12-channel head coil (Siemens Medical Solutions, Erlangen, Germany) at the Berlin Center for Advanced Neuroimaging. In order to reduce motion artifacts, the participant was instructed to fixate on a crosshair and to keep their breath steady for 65 s. Image series were inspected for excessive head movements but no patient exceeded the threshold of 1 mm/TR. After realignment to the first image (including unwarping using the acquired fieldmap) and T1 coregistration onto the mean EPI, rigidly aligned tissue-class images for gray and white matters and cerebrospinal fluid were generated from the coregistered T1 images employing the “New Segment” function. Individual flow fields for warping them to the structural template from 550 healthy adult controls provided with the VBM8 toolbox (Christian Gaser, University of Jena, Germany: http://dbm.neuro.uni-jena.de/vbm/) were created using the DARTEL algorithm. Functional images were then normalized to MNI space and smoothed with a Gaussian kernel of 6 mm FWHM using the normalization function of the DARTEL toolbox. For further analysis, we extracted the gray matter voxels using the respective template contained in SPM8 after binarizing it with a threshold of .5.

2.5. Imaging data preprocessing

Image preprocessing and statistical analyses were carried out using SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK: http://www.fil.ion.ucl.ac.uk/spm/) and Matlab (MathWorks, Natick, MA, USA). Image series were inspected for excessive head movements and no subject exceeded the threshold of 1 mm/TR. After realignment to the first image (including unwarping using the acquired fieldmap) and T1 coregistration onto the mean EPI, rigidly aligned tissue-class images for gray and white matters and cerebrospinal fluid were generated from the coregistered T1 images employing the “New Segment” function. Individual flow fields for warping them to the structural template from 550 healthy adult controls provided with the VBM8 toolbox (Christian Gaser, University of Jena, Germany: http://dbm.neuro.uni-jena.de/vbm/) were created using the DARTEL algorithm. Functional images were then normalized to MNI space and smoothed with a Gaussian kernel of 6 mm FWHM using the normalization function of the DARTEL toolbox. For further analysis, we extracted the gray matter voxels using the respective template contained in SPM8 after binarizing it with a threshold of .5.

2.6. Behavioral data analysis

Two-way repeated-measures ANOVAs with factors “movie” (14 levels) and “condition” (2D versus 3D) were conducted to analyze in-scanner ratings. In case Mauchly’s test indicated a violation of the sphericity assumption, degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity. In addition to post-hoc t-tests (where applicable), two-sided bivariate correlations were calculated to connect in-scanner ratings to values of immersive tendencies as assessed by the QIT. Outlier subjects of more than 2 standard deviations below group means were removed (2 subjects for immersion ratings and 1 subject for the correlation between differential immersion ratings and the QIT).

2.7. fMRI data analysis

In order to find brain networks of activation that are common to all subjects, we introduce the canonical intersubject correlation coefficient (CISC), a multivariate extension of the voxel-wise intersubject correlation measures previously used for analysis of brain activation evoked by complex movie stimuli (Hasson et al., 2004). CISCs are based on canonical correlation analysis (CCA) (Hotelling, 1936) which we used in a similar way as in the analysis of multimodal neuroimaging studies (Biessmann et al., 2011), only that neuroimaging modalities are here replaced by experimental subjects. The underlying assumption is that a network of brain activation for each subject s ∈ {1, 2, ..., S} can be captured as a linear combination of voxels ws ∈ Rs (a V-dimensional vector, where V denotes the number of voxels) of the multivariate voxel time series xS ∈ R²V (T denotes the number of FMR volume). The subscript s indexes the subject and the subscript i indexes the voxel, ws[i] is the ith canonical direction. As the stimulus order was randomized across subjects, fMRI time series needed to be reordered such that each fMRI volume (column of xS) corresponds to the same movie stimulus and frame therein across subjects. Prior to reordering columns of xS, we removed baseline drifts in each voxel time series (rows of xS) by applying a high-pass filter to each row of xS (fifth-order Butterworth filter as implemented in Matlab, cut-off frequency was 0.005 Hz). The linear combinations ws[j] are called canonical directions. We can obtain the time course, also called canonical component, of brain network i for subject s by computing ws'[j]xS. The goal of CISC analysis is to find those canonical directions ws[j] such that the sum over all pairwise correlations (for all pairs of subjects) between the canonical components is maximized, with the constraint that the time courses of two different networks ws[j] and ws[k] be uncorrelated. When concatenating all K canonical directions ws[1], ws[2], ..., ws[K] in a matrix Ws = [ws[1], ws[2], ..., ws[K]] ∈ R²V × K, the objective function of CCA can be formulated as

argmax

w_i, w_j \sum_{j \neq i} Tr(ww^T)X_{ws}[\theta ws[i]w_j]. \forall i, j

subject to \sum_{j \neq i} w_i w_j = I, \forall i.
where \( I \) is the identity matrix. Extensions of classical CCA to sets of variables larger than two are treated in (Kettenring, 1971). The solution of Eq. (1) is given as the top eigenvectors of the generalized eigenvalue equation

\[
\begin{bmatrix}
0 & C_{12} & \ldots & C_{1N} \\
C_{21} & 0 & \ldots & C_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
C_{N1} & C_{N2} & \ldots & 0
\end{bmatrix}
\begin{bmatrix}
W_1 \\
W_2^T(5) \\
\vdots \\
W_N^T(N)
\end{bmatrix}
= \lambda
\begin{bmatrix}
W_1 \\
W_2^T(6) \\
\vdots \\
W_N^T(13)
\end{bmatrix}
\]

(2)

\[Q2\]

Here \( C_{ij} = X_i^T X_j \) denotes the empirical covariance matrix (neglecting normalization constants) between the \( i \)th and \( j \)th subject. For computational efficiency, we computed \( C_{ij} \) not from the full data matrices \( X_i \). We reduced the spatial dimension using principal component analysis (PCA) (Pearson, 1901) and computed \( C_{ij} = \tilde{X}_i^T \tilde{X}_j \) from matrices \( \tilde{X}_i \in \mathbb{R}^{F \times T} \) obtained by projecting the full data matrix onto the \( F \) top eigenvectors \( U_f \) of the spatial covariance matrix \( X_i^T X_i \), so

\[
\tilde{X}_i = U_f^T X_i.
\]

(3)

The principal directions \( U_f \in \mathbb{R}^{F \times T} \) were computed for each subject separately, for details see \( \Lambda \). We kept only as many principal components as were needed to cover 99.9% of the variance in all voxel time series. The number of principal component \( F \) was between 20 and 30, depending on the subject. Importantly, in order to obtain authentic (and not overfitting) estimates of CISCs, we computed the PCA subspaces \( U_f \) and the canonical subspaces \( W_f \) in a leave-one-movie-out cross-validation. For each movie, we estimated \( U_f \) and \( W_f \) on all but this movie (the training data set). The canonical components for the fMRI data recorded during the held-out movie were computed by projecting them onto \( U_f \) and \( W_f \) computed on the training data set.

2.8. Control conditions

We constructed two control conditions based along the lines of standard permutation testing. In a first control condition, we shuffled all fMRI scans in time (complete shuffle condition). All covariations between subjects’ brain activation are removed in this condition. In a second control condition, we tested whether the covariation between brains is movie-specific or reflects merely stimulus-unspecific, generic visual activation. We left the temporal order within each movie block intact and shuffled only the movies’ labels (block shuffle condition). If the intersubject correlations in this condition are as high as in the original unshuffled data set, then the intersubject correlations do not reflect stimulus-specific brain activation, but rather unspecific visual activation.

2.9. Classification of stimulus condition and reported immersion by CISCs

We predicted stimulus condition (2D or 3D) from CISC values in a leave-one-movie-out cross-validation, for more details see Lemm et al. (2011). For each movie, we trained a regularized linear discriminant classifier (LDA) on the CISC values of the most strongly correlated brain networks computed during all but one movie. LDA finds the normal vector \( w_{\text{LDA}} \in \mathbb{R}^F \) of a linear decision boundary by

\[
w_{\text{LDA}} = (S + \lambda I)^{-1}(\mu_+ - \mu_-)
\]

(4)

where \( \mu_+ \) and \( \mu_- \) are the means of the positive and negative class, respectively, \( S \) is the sum of the within-class covariance matrices, and \( \lambda \) is a regularization parameter that is fitted using nested cross-validation within the training data set. The number of networks we used to compute the CISCs, on which the classifier was trained, was \( K = 5 \). We then tested the prediction accuracy and receiver operating characteristic (ROC) of the classifier by predicting the labels on the held-out movie. For the prediction of stimulus conditions, the positive class was the 3D condition and the negative class was data recorded in the 2D condition. For the prediction of immersion reports, we first normalized the immersion ratings by subtracting the mean of each subject’s ratings. We then binarized the normalized ratings and assigned negative labels (low immersion) to normalized ratings smaller than 0 and positive labels (high immersion) to ratings larger than 0. In order to obtain robust estimates of the ROC of the classifier, we performed 100 bootstrap resamplings within the training set.

2.10. Localization of differential CISC strength

The extent to which each voxel reflects a canonical component can be visualized by \( A = W_f^T X_i \) (Haupe et al., 2014). Each column of the matrix \( A_i \in \mathbb{R}^{F \times K} \) contains the spatial pattern of activation corresponding to one canonical component. We compared the patterns of activation, averaged across all movies, between the 2D and the 3D conditions. Patterns were compared in paired t-tests using SPM8. Results were thresholded at \( p = .005 \) and corrected for multiple comparisons (resulting in a whole-brain correction threshold of \( p < .05 \)) by determining individual cluster extent \( k \) thresholds with the calculated intrinsic smoothness of the individual \( F \)-value image, a cluster correction radius of 3 mm, and a 1000-iteration Monte Carlo simulation, using AlphaSim as implemented in the REST Toolbox 1.8 (http://www.restfmri.net/).

Resulting differential patterns, rescaled using SPM8 functions, were associated with the ten most highly correlated psychological concepts per component using the decode function of the online database Neurosynth version 0.3.0 dev (Yarkoni et al., 2011). In an automated and unbiased manner, this function assesses the spatial similarity between an input image and all concept-based meta-analysis maps in its database.

3. Results and discussion

3.1. Behavioral results

In line with the hypothesis that stereoscopic movies are closer to real-world sensory input and enhance the viewer’s engagement with the movie content, a one-sided paired t-test showed significantly stronger immersion of the viewers in the 3D relative to the 2D condition (\( t(22) = 1.91, p = .035 \), Cohen’s \( d = 0.25 \)). The increase in the 3D condition was on average (\( \pm \text{S.E.M.} \) 3.82\% \( \pm 1.77\% \) (see Fig. 4A)).

Differential immersion (3D–2D) showed a significant positive correlation with individual immersive tendencies as reported on the questionnaire for immersive tendencies (QIT) after scanning (Pearson’s \( r(24) = .562, p = .004 \); see Fig. 1). This indicates that the higher the
individual habits to be absorbed in apparent realities, the more strongly
3D movies are experienced as compared to the same movies in 2D.

3.2. 3D movies increase intersubject correlations

In order to assess the effect of stereoscopic stimuli on neural activity, we computed the multivariate canonical intersubject correlation coefficients (CISCs) between all pairs of subjects (see Material and methods) using multiway canonical correlation analysis epholting1936, Kettenring1971.

Fig. 2 shows that in the majority of cases, CISCs were significantly higher when subjects viewed 3D movies than when subjects viewed the same movies in 2D (two-sided paired t-tests; 1st component: t(299) = 6.60, p < .001, 2nd component: t(299) = 6.84, p < .001). CISCs in both conditions were significantly higher than in either control condition (all p < .001), in which we randomly shuffled the movie labels (block shuffle) or all data points (complete shuffle; see Material and methods). For the cortical network that was most correlated across all subjects (the first canonical component), CISCs were .66 ± .010 (mean ± S.E.M.) during 3D movies and .62 ± .010 during 2D movies, while CISCs were .15 ± .007 for the block shuffle condition and .066 ± .003 for the complete shuffle condition. For the second canonical component, CISCs were .61 ± .017 (3D), .56 ± .015 (2D), .15 ± .006 (block shuffle) and .077 ± .003 (complete shuffle).

Fig. 4A shows that the increase in CISCs in the 3D condition was significantly stronger than the increase in immersion as quantified by subjective reports (both p < .001). In the first component the increase was on average 15.0.002,722, while in the second component the increase was 12.3 ± 2.55.

These results confirm and extend the finding from Hasson and colleagues that individual brains “tick collectively” during natural vision (Hasson et al., 2004). Using a bigger sample, different stimuli with varying stereoscopic depth, and multivariate analyses, we show that stereoscopic movie stimuli are associated with increased correlations between brain networks across individuals.

This increase can also be shown for single scenes within a movie. We computed CISCs in sliding windows of 15 second length (or 6 fMRI volumes). Fig. 4C shows the time course of CISC differences between 3D and 2D while subjects were watching a skydiving movie recorded with a head camera. The CISC difference between 3D and 2D was largest in two scenes in which depth cues were pronounced: the jump out of the plane and while landing. During the free fall in-between these two scenes, when there were fewer depth cues, the CISC difference was negligible.

3.3. Visualization and interpretation of CCA networks

In order to visualize and interpret the maximally correlated brain networks, we computed activation maps of their time courses and contrasted the maps of the two conditions (see Material and methods). Fig. 3 shows those brain regions that were significantly more active in the 3D relative to the 2D condition. We did not observe significant activations for the inverse contrast. In order to relate the differential patterns to psychological concepts, we used the decode function of Neurosynth (Yarkoni et al., 2011). Using text-mining techniques in a large corpus of neuroimaging studies, Neurosynth allows to associate contrast patterns with those psychological terms that are most frequently used in studies that report activation in these areas (see Fig. 3). Strongest differential network activations, averaged across movies, were localized in bilateral occipito-temporal regions, which have been associated with visual perception of stereoscopic depth (Rokers et al., 2009), motion, and action (Grosbras et al., 2012). Significant differences were also found in the precuneus and in right-lateralized superior/middle temporal gyrus. Psychological terms associated with these areas describe language- and self-related processes (Cavanna and Trimble, 2006; Tremblay and Small, 2011).

The nonverbal and self-related terms might be connected to the stronger subjective experience of scenes shown in 3D movies. Taken together, these findings provide a basis for future research on how the brain processes dynamic visual information in the real world. Our results suggest that gradually enriching stimuli with more cues in order to bring stimuli closer to real-world experiences should be accompanied by gradual increases in (canonical) intersubject correlations.

3.4. Relationship between intersubject correlations and psychological factors

We investigated whether we can predict from CISCs (a) the stimulus category, and (b) how strongly subjects perceived a movie. We trained a linear classifier on the CISCs in the five most correlated cortical networks estimated during all but one movie. Then, we predicted the stimulus category (2D or 3D) from the CISCs computed on fMRI data recorded while subjects were watching the movie that was excluded from the training set. We found that CISCs can reliably discriminate 3D stimuli from 2D stimuli. The accuracy for decoding the 2D/3D stimulus condition was 64.9% ± 2.9% (mean ± S.E.M. across cross-validation folds) and the bootstrapped area under the receiver operating characteristic curve (AUC) was .70 (see Fig. 4B). Moreover, CISC
Subjective reports of immersion indicate that stereoscopic movies are experienced more strongly than the same movies in 2D. Using multivariate analyses, we find significantly increased intersubject correlations of cortical networks when participants are watching 3D movies. In addition, classifiers trained on intersubject correlations can discriminate 2D from 3D stimuli and high from low immersion ratings. In conclusion, our results highlight the potential of canonical intersubject correlations as a neurophysiological marker not only for visual features such as stereoscopic depth but also for how strongly a stimulus is experienced by human observers.

These findings could have implications for areas beyond basic and cognitive neurosciences. For example, CISCs might be useful for commercial applications and applied research interested in realistic movies. Using CISCs, one could optimize stereoscopic movies by comparing different scenes or presentation techniques and corresponding differences in canonical intersubject correlations. Importantly, CISCs can be used as a non-intrusive marker also on short time scales; this extends their usability beyond behavioral measures or questionnaires, which interfere with a subject’s state of immersion. We have shown that CISC values can be computed on a per-scene basis in sliding windows. If a data base of previously scanned brain activation from other subjects is available, this analysis can also be conducted in an online fashion. The finding that increased intersubject correlations of brain activation is associated with more realistic stimuli and increased immersion may also tie into philosophical debates. For example, it can provide evidence concerning the intersubjective dimension of perception and experience (Gallagher, 2008; Zahavi, 2003) with the potential to connect phenomenological conceptualization and empirical research (Gallagher, 2012).
Finally, our findings could have implications for clinical research. From a clinical perspective, psychopathology entails a patient’s detachment from a reality shared between non-patients. The results of this study suggest that the degree of mental disorders with respect to derealization or detachment from a socially shared reality may be neurophysiologically quantified using intersubject correlations, extending previous findings in autism (Hasson et al., 2009). Ultimately, such quantitative markers could also have the potential to estimate therapeutic success.

Acknowledgments

This work was partially supported by the World Class University Program through the National Research Foundation of Korea funded by the Korean Ministry of Education, Science, and Technology, under grant R31-10008, by the Brain Korea 21 Plus Program through the National Research Foundation of Korea funded by the Ministry of Education, Republic of Korea, by the German Federal Ministry of Education and Research (01GQ0850, 01GQ0411), by the Volkswagen Foundation grant no. II/84051 and by the German Research Foundation (DFG). We thank Judith K. Daniels for the positive influence at early stages of the project and Susanne Herholz as well as Siamac Fazli for valuable comments on an earlier version of this manuscript. J.P.L., M.G., and S.H. conceived and carried out the experiment. F.B. and M.G. analyzed the data. F.B., H.W., K.R.M., M.G., and S.H. wrote the paper. While working on the project F.B. was with TU Berlin and Korea University. The authors declare no conflict of interest.

Appendix A. Row-/column-space eigenvectors

We computed principal component analysis to reduce the spatial dimension (the row-space) of the data matrix $X_s$ for each subject. The principal components $U_i \in \mathbb{R}^n \times F$ are the top $F$ eigenvectors of the covariance matrix of the row-space $X_sX_s^\top \in \mathbb{R}^n \times n$, where $n$ is the number of voxels. As we only have $T$ time samples in our recordings, the estimated covariance matrix $X_sX_s^\top$ has a rank of maximally $T$. As pointed out in Schölkopf et al. (1998), the solution $U_i$ has to lie in a $T$-dimensional subspace of the data matrix; hence the principal directions can be expressed (up to a constant scaling for each eigendirection) as an expansion of the data in $X_s$ as $U_i = X_sR_i$, where $R_i \in \mathbb{R}^n \times T$ are the top $F$ eigenvectors of the Gram matrix $X_sX_s^\top \in \mathbb{R}^n \times T$. The relationship between $U_i$ and $X_s$ can be written as $X_s = U_iR_i^\top$.
between the column space eigenvectors of $X_i X_i$ and the row space eigenvectors of the covariance matrix $X_i X_i$ can be illustrated by considering the singular value decomposition (SVD) of $X_i$

$$X_i = U_i D_i^R \tilde{R}_i^T$$  \tag{A.1}$$

where $U_i$ are orthonormal vectors (i.e. $U_i U_i = I$) forming the PCA basis of the row space, $R_i$ are orthonormal vectors forming the PCA basis of the column space, and $D_i$ is a diagonal matrix containing the singular values. Plugging this SVD approximation into the respective covariance matrices yields

$$X_i X_i' = U_i D_i^R R_i' D_i^R U_i' = U_i D_i^R D_i^R U_i'$$  \tag{A.2}$$

and

$$X_i' X_i = (U_i D_i^R)^T U_i D_i^R = R_i' D_i^R U_i' U_i D_i^R = R_i' D_i^R R_i^T.$$  \tag{A.3}$$

From Eq. (A.1) we see that the eigenvectors of the row-space $U_i$ are

$$U_i = X_i R_i D_i$$  \tag{A.4}$$

and $D_i$ is a diagonal matrix that has $1/D_i(i)$ on its ith diagonal entry and $D_i(i)$ denotes the ith singular value.

References


