NeuroImage 100 (2014) 427–434

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Stereoscopic depth increases intersubject correlations of brain networks

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

[a] Charité – Universitätsmedizin Berlin, Germany
[b] Max Planck Institute for Human Cognitive & Brain Sciences, Leipzig, Germany
[c] Universität Leipzig, Germany
[d] Korea University, Seoul, Republic of Korea
[e] Technische Universität Berlin, Germany
[f] Bernstein Center for Computational Neuroscience, Berlin, Germany
[g] Berlin Center for Advanced Neuroimaging, Berlin, Germany

ARTICLE INFO

Article history:
Accepted 4 June 2014
Available online 17 June 2014

Keywords:
Intersubject correlations
Canonical correlation analysis (CCA)
Natural viewing
3D movies
fMRI
Immersion

ABSTRACT

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 experience. 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. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

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 for human–machine interaction. The addition of stereoscopic depth to visual stimuli aims at approximating everyday-life visual perception, that is, at increasing how realistic a stimulus is. A reduction of sensory differences between artificial stimuli and the physical environment leads to a more engaging viewing experience (Sanchez-Vives and Slater, 2005) — a phenomenon called "immersion" (from Latin: to dive in). Hence, the strength of experience is influenced by aspects of the medium as well as by psycho-physiological processes of the subject. In our study, we assessed both: we measured brain activity with functional magnetic resonance imaging (fMRI) while subjects were freely viewing movies with or without stereoscopic depth. The strength of the experience was acquired through self-report.

Experiments employing complex and dynamic visual stimuli with stereoscopic depth, which are closer to everyday-life visual processing than – especially static – 2D stimuli, can be a first step towards tackling a long-standing and important question behind neuroscientific research: how does our brain process real-world sensory input? To answer this question, neuroscience has extended its focus from studying brain activity of anesthetized animals in response to simple stimuli to studying brain activity of awake brains in response to more complex and realistic stimuli (Maguire, 2012; Zacks et al., 2001). The classical approach using simple stimuli allows to control stimulus parameters systematically in order to study their influence on brain activity (Rust and Movshon, 2005). However, exploring all possible stimuli separately is infeasible. Also, transferring results obtained with the classical approach to real-world scenarios is difficult, as realistic stimuli lead to different brain activation compared to classical simple stimuli (Kayser et al., 2004; Snow et al., 2011). Moreover, neural activity in awake

* Corresponding author at: Machine Learning Group, Technische Universität Berlin, Marchstr. 23, 10587 Berlin, Germany. Fax: +49 30 314 78622.
E-mail address: felix.biessmann@tu-berlin.de (F. Biessmann).
[1] Equal contribution.

http://dx.doi.org/10.1016/j.neuroimage.2014.06.008
1053-8119/© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

animals is very different from brain activity in anesthetized animals (Greenberg et al., 2008).

Therefore, a promising approach to the question how brains process real-world sensory input is to directly study non-anesthetized brain activity in response to complex and more realistic stimuli, for example by non-invasively acquiring neural data during natural vision (Dmochowski et al., 2012; Hasson et al., 2004; Huth et al., 2012).

Analyzing brain responses to complex stimuli under natural viewing conditions is difficult. Classical approaches to analyze neuroimaging data acquired using natural stimuli require considerable human intervention, such as for the design of appropriate filters for feature extraction (Bartels et al., 2008) or labeling of objects and scenes (Huth et al., 2012). In our study, we followed the approach by Hasson and colleagues, who showed that free viewing of complex movie stimuli leads to increased pairwise correlations of single fMRI voxel time courses (Hasson et al., 2004, 2009). We extended their approach by three aspects: first, we used a bigger sample than previous studies. Second, we added stereoscopic depth to movies and quantified associated changes in intersubject correlations. Third, we used multivariate techniques to analyze intersubject correlations of brain *networks* rather than correlations of single voxel time courses. This has two main advantages: multivariate techniques have a higher signal-to-noise ratio (Haufe et al., 2014; Kriegeskorte et al., 2006) and, in contrast to voxelwise correlation analyses, our analysis estimates subject-specific networks and can thus account for inter-individual differences in functional anatomy. We hypothesized that stereoscopic movies are both experienced more strongly and associated with increased correlations of brain activity across movie viewers. Classically, both the dorsal and the ventral pathways of the visual system contribute to the analysis and perception of stereoscopic depth, albeit their neural computations differ (Preston et al., 2008). Especially higher visual regions in middle temporal cortices have been shown to be involved in stereoscopic depth processing both in monkeys (DeAngelis et al., 1998) and in humans (Rokers et al., 2009). Localizing the maximally correlated time courses across movie viewers, we expected the activity pattern to extend beyond the regions in the visual system that are involved in processing stereoscopic information.

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 glasses, 2 by contact lenses, and 16 did not need correction. For those who needed correction, mean dioptric values were −1.75 for the left (*SD*: 2.1, range −5.8 to +2.3) and −1.4 for the right eye (*SD*: 2.1, range −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 naïve 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 $\alpha = .80$), which assessed their personal habits of being drawn into apparent realities like novels or movies (Scheuchenpflug et al., 2003).

Table 1
Description of the movie content.

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

Table 2
Coordinates (in MNI space) and anatomical labels (AAL; Tzourio-Mazoyer et al., 2002) for peaks at which activation maps of time courses of maximally correlated brain networks were significantly higher in the 3D compared to the 2D condition. Calculated using a paired *t*-test on activation patterns thresholded at $p < .005$ with minimal cluster extents *k* that ensured whole-brain correction at $p < .05$; see Fig. 3.

x	y	z	*t* (max)	*p* (uncorrected)	*k* (voxels)	Region (AAL)
Component 1 (k > 108)						
53	−70	−5	5.32	<.001	111	Right inferior temporal
48	−78	−8	3.60	.001		Right inferior occipital
48	−65	0	3.31	.001		Right middle temporal
−40	−63	−5	5.28	<.001	226	Left middle temporal
−40	−65	10	4.72	<.001		Left middle temporal
−53	−68	8	4.44	<.001		Left middle temporal
Component 2 (k > 85)						
13	−53	35	4.85	<.001	107	Right precuneus
8	−63	35	4.02	<.001		Right precuneus
13	−63	45	3.24	.002		Right precuneus
70	−33	3	3.87	<.001	103	Right superior temporal
58	−35	−3	3.74	.001		Right middle temporal
55	−28	−13	3.60	.001		Right middle temporal

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 T_1-weighted image was acquired as a high-resolution anatomical reference using a 3D-MPRAGE sequence with isotropic voxels of 1 mm^3, 192 sagittal slices, repetition time (*TR*) 1900 ms, echo time (*TE*) 2.52 ms, flip angle 9°, field of view (*FoV*) 25 × 25 cm^2. T_2^*-weighted gradient-echo echo-planar images (EPI) were collected for whole-brain functional imaging with isotropic voxels of 2.5 × 2.5 × 2.5 mm^3 (42 axial slices with a 20% distance factor, $TR = 2500$ ms, $TE = 25$ ms, flip angle = 82°, $FoV = 19 \times 19$ cm^2). 476 volumes during viewing were acquired for each participant. Total volume numbers (including fixation periods and rating phases that differed in length for each participant) ranged from 853 to 1042 (mean: 936, *SD*: 43.1). For correcting distortions of the EPI images induced by static magnetic field inhomogeneities, field maps were computed from a standard gradient-echo sequence with two echoes at $TE_1 = 5.19$ ms and $TE_2 = 7.65$ ms, $TR =$ 500 ms, flip angle = 60°, and the geometric parameters matching the EPI images.

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 but no subject exceeded the threshold of 1 mm/*TR*. After realignment to the first image (including unwarping using the acquired fieldmap) and T_1 coregistration onto the mean EPI, rigidly aligned tissue-class images for gray and white matters and cerebrospinal fluid were generated from the coregistered T_1 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 $s \in \{1, 2, \ldots, S\}$ can be captured as a linear combination of voxels $w_{si} \in \mathbb{R}^V$ (a *V*-dimensional vector, where *V* denotes the number of voxels) of the multivariate voxel time series $X_s \in \mathbb{R}^{V \times T}$ (*T* denotes the number of fMRI volumes). The subscript *s* indexes the subject and the subscript *i* indicates that w_{si} is the *i*th canonical direction. As the stimulus order was randomized across subjects, fMRI time series needed to be reordered such that each fMRI volume (column of X_s) corresponds to the same movie stimulus and frame therein across subjects. Prior to reordering columns of X_s, we removed baseline drifts in each voxel time series (rows of X_s) by applying a high-pass filter to each row of X_s (fifth-order Butterworth filter as implemented in Matlab, cut-off frequency was 0.005 Hz). The linear combinations w_{si} are called *canonical directions*. We can obtain the time course, also called *canonical component*, of brain network *i* for subject *s* by computing $w_{si}^\top X_s$. The goal of CISC analysis is to find those canonical directions w_{si} 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 w_{si} and w_{sj} be uncorrelated. When concatenating all *K* canonical directions $w_{s1}, w_{s2}, \ldots, w_{sK}$ in a matrix $W_s = [w_{s1}, w_{s2}, \ldots, w_{sK}] \in \mathbb{R}^{V \times K}$, the objective function of CCA can be formulated as

$$\underset{W_i, W_j}{argmax} \sum_i \sum_j Trace\left(W_i^\top X_i X_j^\top W_j\right), \forall\, i, j \quad \text{subject to } W_i^\top X_i X_i^\top W_i = I, \forall\, i, \tag{1}$$

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

$$\begin{bmatrix} 0 & C_{12} & \dots & C_{1N} \\ C_{21} & 0 & \dots & C_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N1} & C_{N2} & \dots & 0 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_N \end{bmatrix} = \begin{bmatrix} C_{11} & 0 & \dots & 0 \\ 0 & C_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & C_{NN} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_N \end{bmatrix} \Lambda. \tag{2}$$

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

$$\widetilde{X}_s = U_s^\top X_s. \tag{3}$$

The principal directions $U_s \in \mathbb{R}^{V \times F}$ were computed for each subject separately, for details see Appendix A. 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 components F was between 20 and 30, depending on the subject. Importantly, in order to obtain authentic (and not overfitted) estimates of CISCs, we computed the PCA subspaces U_s and the canonical subspaces W_s in a leave-one-movie-out cross-validation. For each movie, we estimated U_s and W_s 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_s and W_s computed on the training data set.

2.8. Control conditions

We constructed two control conditions along the lines of standard permutation testing. In a first control condition, we shuffled all fMRI scans in time (*complete shuffle condition*). All covariation between subjects' brain activations is 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 movie 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 $w_{LDA} \in \mathbb{R}^K$ of a linear decision boundary by

$$w_{LDA} = (S + \lambda I)^{-1}(\mu_+ - \mu_-) \tag{4}$$

where μ_+ and μ_- are the means of the positive and negative class, respectively, S is the sum of the within-class covariance matrices, and λ 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 $A_s = W_s^\top X_s X_s^\top$ (Haufe et al., 2014). Each column of the matrix $A_s \in \mathbb{R}^{V \times K}$ 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 T-value image, a cluster connection 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, resized 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 automatized 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 (±S.E.M.) 3.82% ± 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

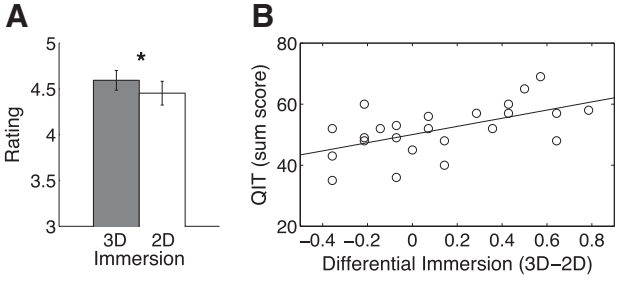

Fig. 1. (A) Participants experienced 3D movies more strongly (immersion; 1: weak, 7: strong) than the same movies in 2D (one-sided post-hoc t-test; $t(22) = 1.91$, $p = .035$, $d = 0.25$). Error bars: S.E.M. (B) Differential immersion (3D–2D) was positively correlated with individual immersive tendencies, as measured by a questionnaire (QIT) after scanning (Pearson's $r(24) = .562$, $p = .004$).

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 (Hotelling, 1936; Kettenring, 1971).

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).

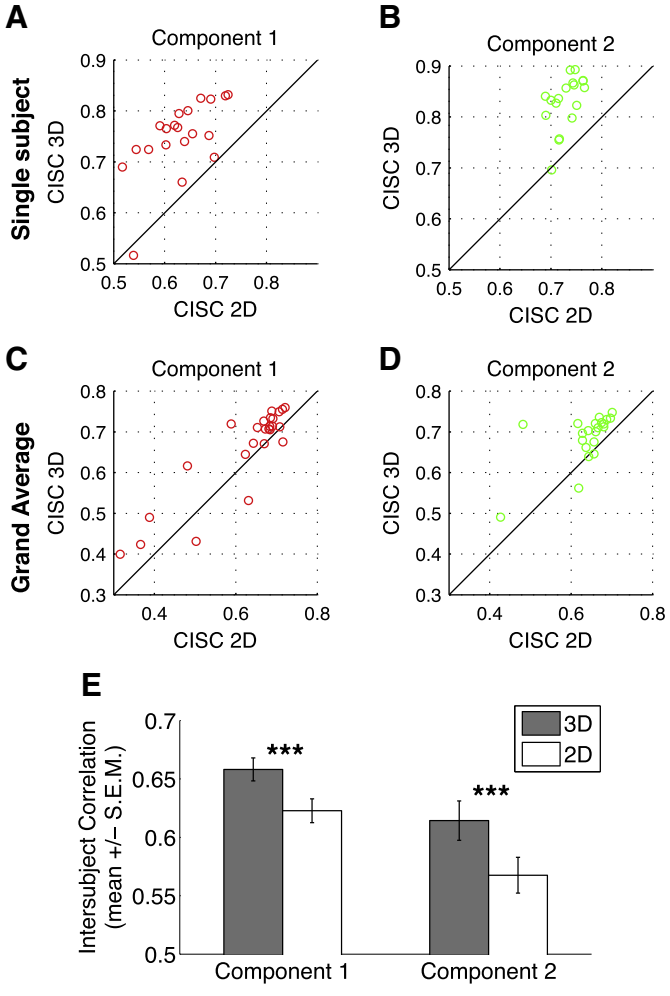

Fig. 2. Stereoscopic depth increased canonical intersubject correlations on both the single-subject and the group level. (A to D) Scatter plots showing canonical intersubject correlation coefficients (CISCs) for 2D (x-axis) and 3D (y-axis). Circles above the diagonal indicate higher correlations in the 3D condition. (A and B) CISCs between a single subject and all other subjects for the first (A) and the second (B) canonical components. (C and D) CISCs for each subject averaged across all subject pairs for the first (C) and the second (D) canonical components. (E) Mean CISCs (±S.E.M.) in canonical components 1 and 2 during 3D compared to 2D viewing.

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 components $p < .001$). In the first component the increase was on average 15.0% ± 2.72%, 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

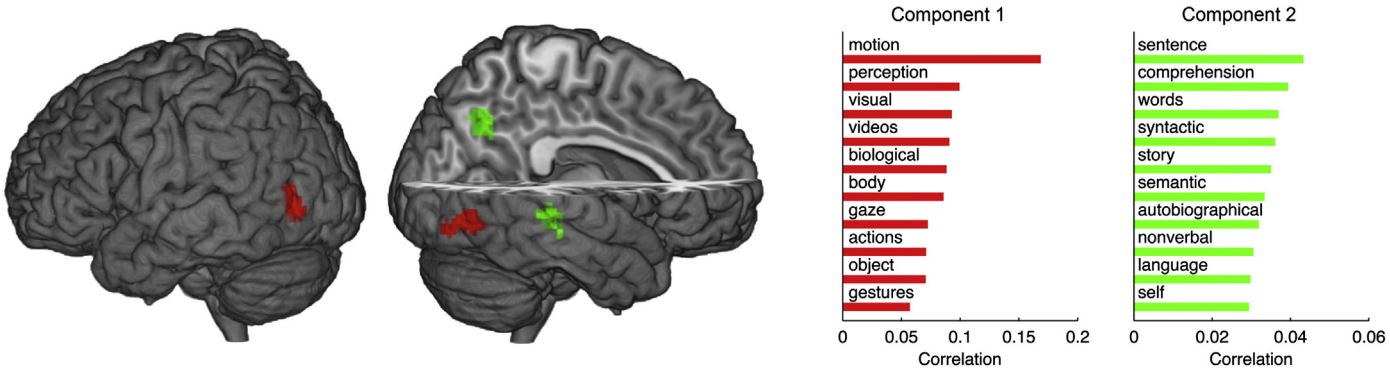

Fig. 3. Activation maps of time courses of maximally correlated brain networks were contrasted (3D–2D) and thresholded at $p < .005$ with individual cluster extent thresholds for whole-brain correction at $p < .05$ (component 1: $k > 108$, component 2: $k > 85$). For exact coordinates see Table 2. Patterns in component 1 (red) and component 2 (green) were associated with psychological concepts using the *decode* function of Neurosynth (Yarkoni et al., 2011). The x-axes represent correlation values between the contrast pattern and each concept's meta-analysis map.

values could discriminate high and low immersion ratings of movies (accuracy 54.9% ± 2.0%, AUC = .56). Although the classification accuracy of immersion ratings on hold-out data was rather low, it was significantly higher than chance level, which was at 50% (two-sided one-sample *t*-test; $t(13) = 2.42$, $p = .031$), indicating that it is possible to infer from the strength of intersubject correlations whether or not a subject experiences a movie as engaging.

Note that the decoding accuracy of stimulus category (2D/3D) was significantly higher than that of subjective reports of immersion (two-sided paired *t*-test; $t(13) = 2.79$, $p = .015$). Likewise, the increase in subjective immersion ratings was smaller than the increase in CISC values in the 3D condition (see Fig. 4C). A possible explanation for these differences is that the strength of an experience is strongly influenced by factors other than stereoscopic depth, such as content and narrative structure of a movie or the possibility to interact with the medium. In our study, adding "interaction" as experimental factor was not only difficult given the chosen stimuli and limitations imposed by the MRI environment. Another difficulty with interactive paradigms in our setting is that our analysis requires all subjects to have seen the same stimulus in order to calculate intersubjective correlations. Furthermore, by presenting all movie stimuli (i.e., each movie twice) to each subject and in randomized order, we deliberately eliminated subjective and complex movie features related to content and narrative structures. This enabled us to isolate the influence of stereoscopic stimuli on immersion and on CISCs. Evaluating this effect, we found that the increase in CISC values was substantially larger than the increase in immersion. In line with previous studies, this difference suggests that for quantification of how strongly a complex and dynamic stimulus is experienced, neurophysiological markers are a valuable complement to subjective reports (Porbadnigk et al., 2013; Scholler et al., 2012).

4. Conclusions

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 are associated with more realistic stimuli and increased immersion

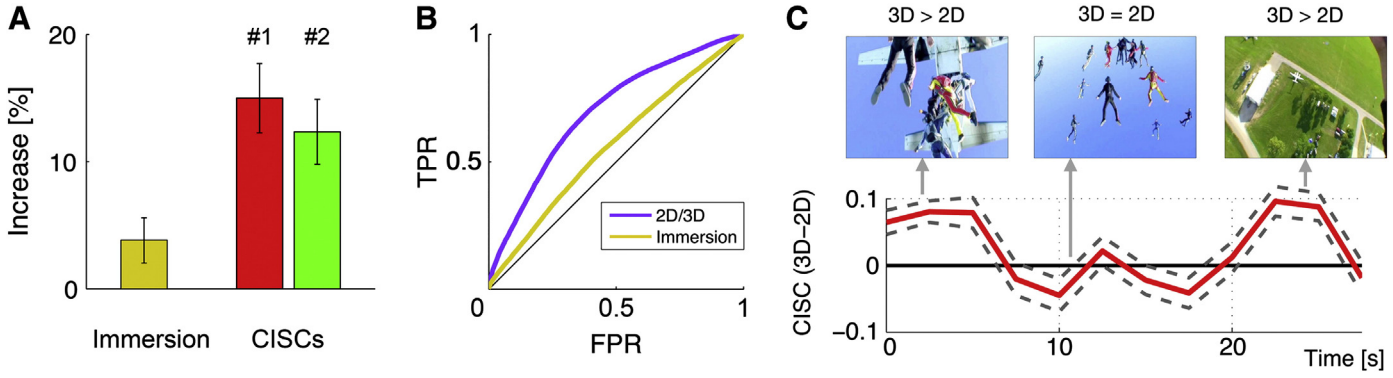

Fig. 4. (A) Stereoscopic movie stimuli result in a moderate increase of reported immersion and a strong increase in canonical intersubject correlations (CISCs) in components 1 and 2. (B) CISCs discriminate 2D from 3D stimuli on hold-out data (blue line) and discriminate high and low immersion ratings (brown line). The receiver operating characteristic (ROC) curves show the true positive rate (TPR) of a linear classifier as a function of its false positive rate (FPR). (C) Difference of CISCs (3D–2D condition) in the first canonical component during a skydiving movie (mean ± S.E.M.), computed in time windows of 15 s. Note the strong increase in CISCs in the 3D condition when the camera man jumps out of the plane and when he is landing.

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 $U_s \in \mathbb{R}^{V \times F}$ are the top F eigenvectors of the covariance matrix of the row-space $X_s X_s^\top \in \mathbb{R}^{V \times V}$, where V is the number of voxels. As we only have T time samples in our recordings, the estimated covariance matrix $X_s X_s^\top$ has a rank of maximally T. As pointed out in Schölkopf et al. (1998), the solution U_s 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 $U_s = X_s R_s$, where $R_s \in \mathbb{R}^{F \times T}$ are the top F eigenvectors of the Gram matrix $X_s^\top X_s \in \mathbb{R}^{T \times T}$. The relationship between the column space eigenvectors of $X_s^\top X_s$ and the row space eigenvectors of the covariance matrix $X_s X_s^\top$ can be illustrated by considering the singular value decomposition (SVD) of X_s

$$X_s = U_s D_s R_s^\top \tag{A.1}$$

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

$$\begin{aligned} X_s X_s^\top &= U_s D_s R_s^\top \left(U_s D_s R_s^\top\right)^\top \\ &= U_s D_s \underbrace{R_s^\top R_s}_{I} D_s^\top U_s^\top = U_s D_s^2 U_s^\top \end{aligned} \tag{A.2}$$

and

$$\begin{aligned} X_s^\top X_s &= \left(U_s D_s R_s^\top\right)^\top U_s D_s R_s^\top \\ &= R_s D_s^\top \underbrace{U_s^\top U_s}_{I} D_s R_s^\top = R_s D_s^2 R_s^\top . \end{aligned} \tag{A.3}$$

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

$$U_s = X_s R_s \hat{D}_s \tag{A.4}$$

where $\hat{D}_s$ is a diagonal matrix that has $1/D_{s(i,i)}$ on its ith diagonal entry and $D_{s(i,i)}$ denotes the ith singular value.

References

Bartels, A., Zeki, S., Logothetis, N.K., 2008. Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. Cereb. Cortex 18 (3), 705 (Mar).

Biessmann, F., Plis, S.M., Meinecke, F.C., Eichele, T., Müller, K.-R., 2011. Analysis of multimodal neuroimaging data. IEEE Rev. Biomed. Eng. 4, 26–58.

Cavanna, A.E., Trimble, M.R., 2006. The precuneus: a review of its functional anatomy and behavioural correlates. Brain 129 (3), 564–583. http://dx.doi.org/10.1093/brain/awl004.

Crone, R.A., 1992. The history of stereoscopy. Doc. Ophthalmol. 81 (1), 1–16 (Jan. URL http://www.ncbi.nlm.nih.gov/pubmed/1473457).

DeAngelis, G.C., Cumming, B.G., Newsome, W., 1998. Cortical area MT and the perception of stereoscopic depth. Nature 394, 677–680. http://dx.doi.org/10.1038/29299.

Dmochowski, J.P., Sajda, P., Dias, J., Parra, L.C., 2012. Components of ongoing eeg with high correlation point to emotionally-laden attention — a possible marker of engagement? Front. Human Neurosci. 6 (112). http://dx.doi.org/10.3389/fnhum.2012.00112.

Gallagher, S., 2008. Intersubjectivity in perception. Cont. Philos. Rev. 41 (2), 163–178. http://dx.doi.org/10.1007/s11007-008-9075-8 (Jul).

Gallagher, S., 2012. On the possibility of naturalizing phenomenology. In: Zahavi, D. (Ed.), Oxford Handbook of Contemporary Phenomenology. 4. Oxford University Press, Oxford, Ch, pp. 70–93.

Greenberg, D.S., Houweling, A.R., Kerr, J.N.D., 2008. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. Nat. Neurosci. 11 (7), 749–751. http://dx.doi.org/10.1038/nn.2140 (Jul).

Grosbras, M.-H., Beaton, S., Eickhoff, S.B., 2012. Brain regions involved in human movement perception: a quantitative voxel-based meta-analysis. Hum. Brain Mapp. 33 (2), 431–454. http://dx.doi.org/10.1002/hbm.21222.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303 (5664), 1634–1640. http://dx.doi.org/10.1126/science.1089506 (Mar).

Hasson, U., Avidan, G., Gelbard, H., Vallines, I., Harel, M., Minshew, N., Behrmann, M., 2009. Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. Autism Res. 2 (4), 220–231. http://dx.doi.org/10.1002/aur.89.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., Biessmann, F., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage 87, 96–110. http://dx.doi.org/10.1016/j.neuroimage.2013.10.067.

Hotelling, H., 1936. Relations between two sets of variates. Biometrika 28 (3), 321–377. http://dx.doi.org/10.1093/biomet/28.3-4.321.

Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76 (6), 1210–1224. http://dx.doi.org/10.1016/j.neuron.2012.10.014 (Dec).

Kayser, C., Kim, M., Ugurbil, K., Kim, D.-S., König, P., 2004. A comparison of hemodynamic and neural responses in cat visual cortex using complex stimuli. Cereb. Cortex 14 (8), 881–891. http://dx.doi.org/10.1093/cercor/bhh047 (Aug).

Kettenring, J.R., 1971. Canonical analysis of several sets of variables. Biometrika 58 (3), 433–451. http://dx.doi.org/10.1093/biomet/58.3.433.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. 103 (10), 3863. http://dx.doi.org/10.1073/pnas.0600244103.

Lemm, S., Blankertz, B., Dickhaus, T., Müller, K.-R., 2011. Introduction to machine learning for brain imaging. Neuroimage 56 (2), 387–399. http://dx.doi.org/10.1016/j.neuroimage.2010.11.004 (May).

Maguire, E.A., 2012. Studying the freely-behaving brain with fMRI. NeuroImage 62 (2), 1170–1176. http://dx.doi.org/10.1016/j.neuroimage.2012.01.009 (Aug).

Oldfield, R., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97–113. http://dx.doi.org/10.1016/0028-3932(71)90067-4 (Mar).

Parker, A.J., 2007. Binocular depth perception and the cerebral cortex. Nat. Rev. Neurosci. 8 (5), 379–391. http://dx.doi.org/10.1038/nrn2131 May.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Philos. Mag. 2, 559–572.

Porbadnigk, A.K., Treder, M.S., Blankertz, B., Antons, J.-N., Schleicher, R., Möller, S., Curio, G., Müller, K.-R., 2013. Single-trial analysis of the neural correlates of speech quality

perception. J. Neural Eng. 10 (5), 056003. http://dx.doi.org/10.1088/1741-2560/10/5/056003 (Oct).
Preston, T.J., Li, S., Kourtzi, Z., Welchman, A.E., 2008. Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. J. Neurosci. 28 (44), 11315–11327. http://dx.doi.org/10.1523/JNEUROSCI.2728-08.2008 (Oct).
Rokers, B., Cormack, L.K., Huk, A.C., 2009. Disparity- and velocity-based signals for three-dimensional motion perception in human MT+. Nat. Neurosci. 12 (8), 1050–1055. http://dx.doi.org/10.1038/nn.2343 (Aug).
Rust, N.C., Movshon, J.A., 2005. In praise of artifice. Nat. Neurosci. 8 (12), 1647–1650. http://dx.doi.org/10.1038/nn1606 (Dec).
Sanchez-Vives, M.V., Slater, M., 2005. From presence to consciousness through virtual reality. Nat. Rev. Neurosci. 6 (4), 332–339. http://dx.doi.org/10.1038/nrn1651 (Apr).
Scheuchenpflug, R., Ruspa, C., Quattrocolo, S., 2003. Presence in virtual driving simulators. In: de Waard, D., Brookshuis, K., Breker, S., Verwey, W. (Eds.), Human Factors in the Age of Virtual Reality. Shaker Publishing, Maastricht, NL, pp. 143–148.
Schölkopf, B., Smola, A.J., Müller, K.-R., 1998. Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. 10 (6), 1299–1319. http://dx.doi.org/10.1162/089976698300017467.
Scholler, S., Bosse, S., Treder, M.S., Blankertz, B., Curio, G., Müller, K.-R., Wiegand, T., 2012. Toward a direct measure of video quality perception using eeg. IEEE Trans. Image Process. 21 (5), 2619–2629. http://dx.doi.org/10.1109/TIP.2012.2187672 (May).
Snow, J.C., Pettypiece, C.E., McAdam, T.D., McLean, A.D., Stroman, P.W., Goodale, M. a, Culham, J.C., 2011. Bringing the real world into the fMRI scanner: repetition effects for pictures versus real objects. Sci. Rep. 1, 130. http://dx.doi.org/10.1038/srep00130 (Jan).
Tremblay, P., Small, S.L., 2011. From language comprehension to action understanding and back again. Cereb. Cortex 21 (5), 1166–1177. http://dx.doi.org/10.1093/cercor/bhq189.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15 (1), 273–289. http://dx.doi.org/10.1006/nimg.2001.0978 (Jan).
Wheatstone, C., 1838. Contributions to the physiology of vision. part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. Philos. Trans. R. Soc. Lond. 128, 371–394 (URL http://rstl.royalsocietypublishing.org/content/128/371.short).
Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8 (8), 665–670. http://dx.doi.org/10.1038/nmeth.1635 (Aug).
Zacks, J., Braver, T., Sheridan, M., 2001. Human brain activity time-locked to perceptual event boundaries. Nat. Neurosci. 4, 651–655. http://dx.doi.org/10.1038/88486.
Zahavi, D., 2003. Husserl's Phenomenology. Stanford University Press, Stanford.