bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Psychophysiological Interactions in a Visual Checkerboard Task: Reproducibility, Reliability, and the Effects of Deconvolution

Xin Di and Bharat B. Biswal*

Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ, 07102, USA.

*Corresponding author:

Bharat B. Biswal, PhD

607 Fenster Hall, University Height

Newark, NJ, 07102, USA

bbiswal@yahoo.com

Running head: Reproducibility and reliability of PPI

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract:

Psychophysiological interaction (PPI) is a regression based method to study task modulated brain connectivity. Despite its popularity in functional MRI (fMRI) studies, its reliability and reproducibility have not been evaluated. We investigated reproducibility and reliability of PPI effects during a simple visual task, and examined the effect of deconvolution on the PPI results. A large open-access dataset was analyzed (n = 138), where a visual task was scanned twice with repetition times (TRs) of 645 ms and 1400 ms, respectively. We first replicated previous results by using the left and right middle occipital gyrus as seeds. Then ROI-wise (regions of interest) analysis was performed among twenty visual-related thalamic and cortical regions, and negative PPI effects were found between many ROIs with the posterior fusiform gyrus as a hub region. Both the seed-based and ROI-wise results were similar between the two runs and between the two PPI methods with and without deconvolution. The non-deconvolution method and the short TR run in general had larger effect size and greater extent. However, the deconvolution method performed worse in the 645 ms TR run than the 1400 ms TR run in the voxel-wise analysis. Given the general similar results between the two methods and the uncertainty of deconvolution, we suggest that deconvolution may be not necessary for PPI analysis on block-designed data. Lastly, intraclass correlations between the two runs were much lower for PPI effects than activation main effects, which raise cautions on performing inter-subject correlations and group comparisons on PPI effects.

Keywords: Reproducibility, reliability, test-retest, psychophysiological interaction, deconvolution.

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. Introduction

Psychophysiological interaction (PPI) is a widely used method to study task related brain functional connectivity changes (Friston et al., 1997). It employed simple regression-based method to model task modulated connectivity effects, thus enabling whole brain exploratory analysis. Therefore, even though there are more sophisticated methods available, e.g. dynamic causal modeling (Friston et al., 2003), PPI is still a valuable method for fMRI data, given that our knowledge on large-scale task related connectivity is still quite limited.

A PPI effect is defined as an interaction between the time series of a brain region (physiological variable) and a (or more) task design variable (psychological variable). Noises of both the physiological and psychological variables go into the interaction term, so that the interaction effect is much noisier than the main effects of task free connectivity (physiological main effect) and task activation (psychological main effect). This makes PPI analysis having lower statistical power than simple connectivity and conventional activation analysis. Since PPI analysis has been increasingly used to study group differences and inter-subjects variability, it is important to evaluate the reproducibility and reliability of the PPI methods (Dubois and Adolphs, 2016; Vul et al., 2009). Voxel-based meta-analysis has been used to examine consistency of PPI results across studies (Di et al., 2017a). However, because the tasks used in different studies varied greatly, the motivation of a meta-analysis on PPI was rather to identify different connectivity that were modulated by different tasks, than to simply identify consistent connectivity cross studies with different tasks (Di et al., 2017a). Nevertheless, the reliability of PPI effect has not been directly examined.

One critical step for the PPI method is to properly deal with the asynchrony between task design and observed blood-oxygen-level dependent (BOLD) signals. An earlier solution is to convolve the psychological variable with hemodynamic response function (HRF). Then the PPI term x^1_{PPI} could be expressed as:

$$x^1_{PPI} = x_{Physio} \cdot (z_{Psych} * hrf) \qquad (1)$$

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where x_{Physio} represents the physiological variable, z_{Psych} represents the psychological design variable, and * represents convolution operator. However, this calculation is not appropriate if the interaction happened faster than the slow hemodynamic response. Therefore, a deconvolution procedure is required (Gitelman et al., 2003) to find a variable z_{Physio} that:

$$x_{Physio} = z_{Physioi} * hrf \qquad (2)$$

If this could be achieved, then the interaction could be calculated at the neuronal level and then convolve with HRF:

$$x^2_{PPI} = (z_{Psych} \cdot z_{Physio}) * hrf \qquad (3)$$

We can also put equation (2) to equation (1), so that:

$$x^1_{PPI} = (z_{Psych} * hrf) \cdot (z_{Physio} * hrf) \qquad (4)$$

Mathematically, x^1_{PPI} and x^2_{PPI} are not equivalent. Therefore, deconvolution seems necessary. However, deconvolution is not an easy task, with assumptions such as known HRF and noise characteristics in the BOLD signals (O'Reilly et al., 2012; Roebroeck et al., 2011). Unfortunately, there are substantial amount of variability in HRF both across brain regions and across subjects (Handwerker et al., 2004). On the other hand, with a known physiological variable at neuronal level, we demonstrate by simulation that the PPI terms calculated from the two methods (x^1_{PPI} and x^2_{PPI}) are actually quite similar in a typical block-design (Figure 1). In this simulation, we defined a simple block-designed task with equal on and off periods with different cycle lengths (from 8 s to 80 s), and a simple event-related design with fixed inter trial interval of 12 s. We then generated the physiological variable as a Gaussian variable for 1000 times. As can be seen in Figure 1, for a block design, with a typical design cycle (greater than 40 s, i.e. 20-s on and 20-s off), the correlation of PPI terms could be higher than 0.9. For the event-related design and the fast block-design (8 s cycle, i.e. 4 s-on and 4-s off), however, the mean correlation are below 0.5 with large variations. Whether to perform deconvolution then needs to balance between the deviation between the PPI terms calculated in different ways and the uncertainty of deconvolution (Di et al., 2017b). Therefore, it might be better to not perform deconvolution for an block-designed task, which

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is actually recommended by FSL (FMRIB Software Library) (Jenkinson et al., 2012; O'Reilly et al., 2012). We note that deconvolution is still necessary for an event-related task, but a recent study has suggested that beta series method might be a better alternative (Cisler et al., 2014).

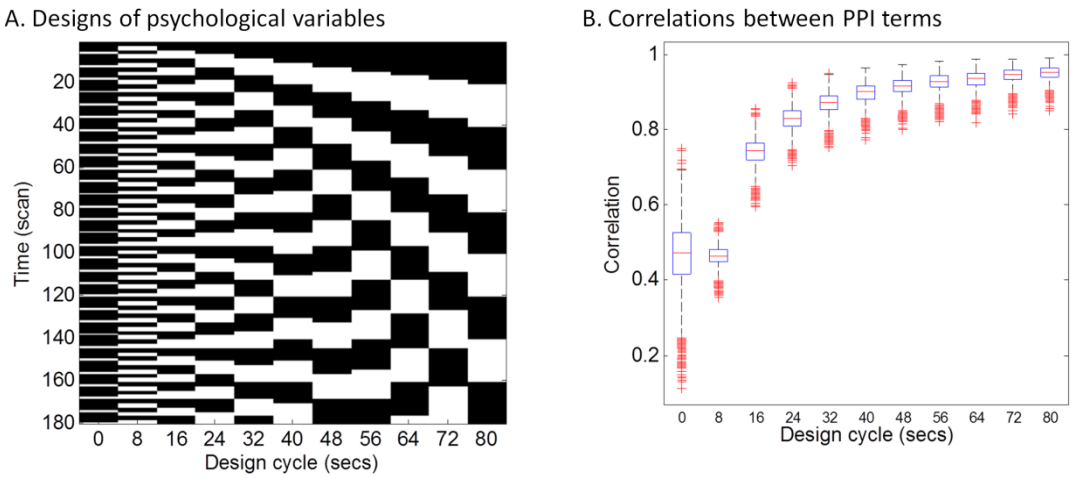

Figure 1 Simulations of the correlations between PPI terms calculated from deconvolution and non-deconvolution methods. Panel A illustrates different task designs that were used for the simulation. Each column represents a task design. 0 of x axis represents the event-related design, with 1 time bin (2 s) of trial condition and 5 time bins (10 s) of baseline condition. The remaining columns show block designs with different frequencies of repetition. For example, 80 secs cycle means 40-s on and 40-s off of the task condition related to baseline. A physiological variable was generated using Gaussian random noise for 1,000 times. Penal B shows boxplots of correlations between PPI terms calculated from deconvolution and non-deconvolution methods with the 1,000 simulations. The code for this simulation can be found at: https://github.com/dixy0/PPI_correlation_demo .

We recently demonstrated negative PPI effects (reduced connectivity) between the middle occipital gyrus to fusiform gyrus and supplementary motor areas in a simple block-designed checkerboard task (Di et al., 2017b). Here, we further analyzed a larger sample of checkerboard data (n = 138) of two separate runs with two repetition times (TR: 645 ms and 1400 ms) (Nooner et al., 2012). The aims of the current study are to first evaluate reproducibility and reliability of PPI effects in the checkerboard task.

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Additionally, we investigated the impact of PPI calculation methods on the PPI results and their reproducibility and reliability. We operationally defined reproducibility as whether previously reported clusters could be observed in the current analysis, and whether the clusters reported in one run could be observed in the other run. Quantitatively, we utilized Dice coefficient to quantify overlaps of voxels on thresholded maps (Rombouts et al., 1998; Taylor et al., 2012). Next, we used intraclass correlation (ICC) to quantify test-retest reliability. Because the short TR run has about twice the number of time points as the long TR one, we predict that statistical results would be better for the short TR run compared with the long TR run. In addition, shorter sampling rate may provide more accurate estimate of hemodynamic response, therefore deconvolution PPI method should work better for the short TR than the long TR runs.

2. Methods

2.1 Tasks and fMRI data

We used the checkerboard fMRI data with TRs of 645 ms and 1400 ms from the release 1 of Enhanced Nathan Kline Institute - Rockland Sample (http://fcon_1000.projects.nitrc.org/indi/enhanced/). 146 subjects' data with age equal or larger than 20 years old were included for analysis. Six subjects' data were discarded due to large head motion during fMRI scanning in any of the two scans (maximum frame-wise displacement (FD) (Di and Biswal, 2015) greater than 1.5 mm or 1.5^{o}). One subject's data were deleted because of poor coverage of the lower occipital lobe, and another subject's data were deleted because of failure of coregistration and normalization. The effective number of subjects was 138 (89 females, 45 males, 1 unidentified). The mean age of the sample was 47.8 years (20 to 83 years).

The checkerboard task consisted of 20 s fixation block and 20 s flickering checkerboard block repeated three times. A blank screen was presented after the third checkerboard block until fMRI scan was complete. The task was scanned for two separate runs with two TRs: 645 ms and 1400 ms, respectively. For the 645 ms run, 239 or 240 fMRI images were scanned for each subject. The following parameters were used: TR = 645 ms; TE = 30 ms; flip angle = 60 deg; voxel size = 3 x 3 x 3 mm^3

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

isotropic; number of slices = 40. For the 1400 ms run, 98 fMRI images were scanned for each subject. The following parameters were used: TR = 1400 ms; TE = 30 ms; flip angle = 65 deg; voxel size = 2 x 2 x 2 mm^3 isotropic; number of slices = 64. Anatomical T1 images were scanned using MPRAGE (magnetization-prepared rapid acquisition with gradient echo) sequence with the following parameters: TR = 1900 ms; TE = 2.52 ms; flip angle = 9 ^{o}; voxel size = 1 x 1 x 1 mm^3 isotropic. More information of the data can be found in Nooner et al. (Nooner et al., 2012).

2.2 FMRI data analysis

2.2.1 FMRI data preprocessing

Functional MRI data preprocessing and analysis were performed using SPM12 software (http://www.fil.ion.ucl.ac.uk/spm/) under MATLAB environment (http://www.mathworks.com/). For the 645 ms run, the first 14 images (9 s) were discarded from analysis, resulting in 225 images for each subject. For the 1400 TR run, the first five images (7 s) were discarded from analysis, resulting in 93 images for each subject. The functional images were motion corrected, and corregistered to subject's anatomical images. The anatomical images were segmented, and the deformation field images were used to normalize the functional images into MNI space. The data from the two TR runs were both resliced and resampled at a spatial resolution of 3 x 3 x 3 mm^3. Lastly, the functional images were smoothed using a 6 mm full width at half maximum (FWHM) Gaussian kernel.

2.2.2 Activation analysis

We first defined functional ROIs of the visual thalamus and lower visual area by performing general linear model (GLM) analysis on the checkerboard task. The checkerboard task was modeled as a box-car function, and was convolved with canonical hemodynamic response function (HRF) to form a predictor of BOLD responses. Two regressors of the first eigenvariate of BOLD signals in white matter and cerebrospinal fluid (CSF), and 24 regressors of Friston's autoregressive head motion model (Friston et al., 1996) were also added in the model as covariates. An implicit high-pass filter of 1/128 Hz was also implemented in the model. The GLM model was estimated for each voxel in the brain to identify regions

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that showed similar patterns of activations as the task design. The beta maps of task activation were used for group level analysis using a one sample t-test model. Statistical significant clusters were identified by using cluster level statistics based on random field theory. Clusters were first identified using a one-tailed t-test at $p < 0.001$, and cluster extent was determined using false discovery rate (FDR) at $p < 0.05$.

2.2.3. Definition of regions of interest

We performed two types of PPI analyses, voxel-wise analysis using seed regions that were activated by the checkerboard task and ROI-based analysis among visual thalamus and cortical visual areas independently defined from other toolbox. In the activation analysis of the current data, the posterior visual cortex and the posterior portion of the thalamus were robustly activated by the visual checkerboard stimulation in both TR runs. We therefore defined the left and right middle occipital gyrus (LMOG and RMOG) and the thalamus as regions of interest (ROIs) based on the activations. To define the ROIs with proper size, we increase the threshold to $t > 16$ to define the LMOG and RMOG, and made an intersection between the two runs. The size of LMOG was 222 voxels, and the size of RMOG was 259 voxels. Thalamus was defined using a threshold of $p < 0.001$, with an intersection between the two runs. Because the visual thalamus is small, left and right ROIs were combined to form a single thalamus ROI (171 voxels). Different thresholds were chosen to ensure that these ROIs are similar in size. The eigenvariate of a ROI was extracted with adjustment of effects of no interests (head motion, WM/CSF variables, and low frequency drifts).

We defined the visual thalamus as the regions that show functional associations with the lateral visual network in resting-state (Yuan et al., 2016). Cortical visual areas were defined by using probabilistic cytoarchitectonic maps. These areas include the OC1/OC2 (occipital cortex) (Amunts et al., 2000), ventral and dorsal OC3 and OC4 (Kujovic et al., 2013; Rottschy et al., 2007), OC5 (Malikovic et al., 2006), and FG1/FG2 (fusiform gyrus) (Caspers et al., 2013). For the probabilistic maps of these regions, we first performed a winner-takes-all algorithm to define unique regions of each area, and then split them into left and right regions. As a result, there are 20 ROIs (left and right thalamus, OC1, OC3,

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

OC3d, OC3v, OC4d, OC4v, OC5, FG1, and FG2). The eigenvariate of a ROI was extracted with adjustment of effects of no interests (head motion, WM/CSF variables, and low frequency drifts).

2.2.4. Psychophysiological interaction analysis

PPI analysis was performed using SPM12 with updates 6685. PPI terms were calculated by using both deconvolution method and non-deconvolution method. For the deconvolution method, the time series of a seed region was deconvolved with the canonical HRF, multiplied with the centered psychological box-car function, and convolved back with the HRF to form a predicted PPI time series at hemodynamic response level. For the non-deconvolution method, the box-car function of psychological design was convolved with the HRF to form a psychological variable, and it was centered and multiplied with the raw seed time series. Figure 2 shows examples of PPI terms calculated from the two methods in the two TR runs.

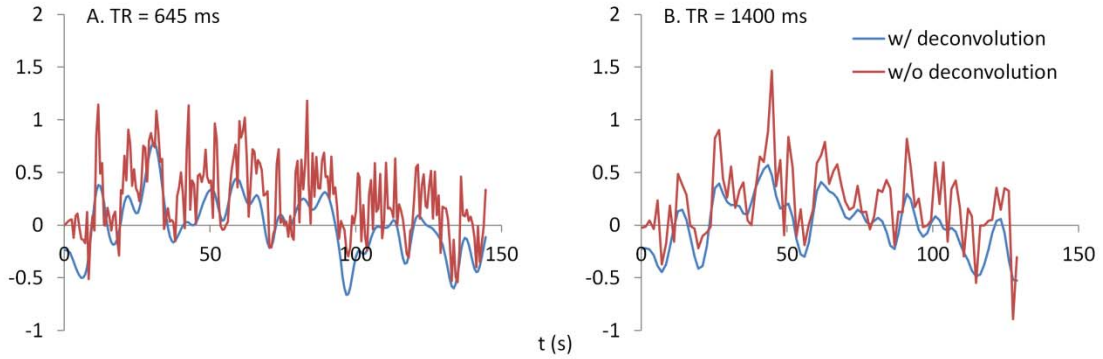

Figure 2 Examples of PPI terms calculated by deconvolution and non-deconvolution methods for the two TR runs.

For voxel-wise PPI analysis, separate GLMs were built for the LMOG, RMOG, and thalamus seeds, and for the two TR runs. The models included one regressor representing task activation, one regressor representing the seed time series, the PPI term, and covariates the same as the activation GLMs descripted above. Group-level one sample t-test was used on the corresponding PPI effects, to test where in the brain showed consistent PPI effects with a seed region. For both positive and negative contrasts, a

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

one-tailed t-test of $p < 0.001$ was first used to define clusters, and then a FDR cluster threshold of $p< 0.05$ was used to identify statistical significant clusters. For ROI-wise analysis, PPI GLM models were built for each of the 20 ROIs, and applied to all other ROIs as a dependent variable. The GLM model included one psychological variable, one physiological variable, one PPI variable, and one constant term. The covariates were not included because they have already been regressed out from all ROI time series. PPI effects were calculated between each pair of ROIs, resulting in a 20 x 20 matrix of beta values for each subject. The matrices were symmetrized by averaging corresponding upper and lower diagonal elements (Di et al., 2017b), with a total of 190 (20 x 19 / 2) unique effects. Group-level one-sample t-test was performed on each element of the matrix. For both positive and negative contrasts, a one-tailed t-test of $p < 0.001$ was used to identify significant PPI effects. This threshold was chosen to match with voxel-wise analysis. We also used FDR correction on the total of 190 effects. And the results are similar to what using a $p < 0.001$ threshold. However, FDR depends on the distribution of all tested p values, making it difficult to compare between two runs. Therefore, we adopted $p < 0.001$ to report ROI-based PPI results.

2.2.5. Reproducibility and reliability

We operationally define reproducibility as overlaps of super-threshold clusters. Dice coefficient was used to quantify reproducibility (Rombouts et al., 1998). Two strategies were used to threshold the maps or matrix from the two TR runs. First, statistical t maps or t matrices from the two TR runs were thresholded using a common t value, ranging from 1.7 (approximately corresponds to $p < 0.05$) to 7. However, it is possible that the effect sizes in the two TR runs are systematically different, so that using a same t value could generate very different numbers of superthreshold voxels or elements in the two runs. Therefore, we also thresholded t maps or t matrices based on the percentile of t values within a map or matrix. This could ensure that the numbers of superthreshold voxels or elements are the same between the two TR runs.

We operationally define reliability as test-retest reliability between the two TR runs, as quantified as ICC (Zuo et al., 2010a). Voxel-wise ICC maps or each ROI and ICC matrices across 20 ROIs were

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

calculated between two TR runs for each PPI method. At each voxel or matrix element, ICC was calculated from a 138 (subject) by 2 (run) matrix by using a MATLAB function written by Zuo et al. (Zuo et al., 2010a). Because only voxels that have significant effects might show meaningful reliability, we displayed histograms of ICCs within significant voxels or elements with reference to those in the whole brain. For task activations, the significant voxels were determined using intersection of the two TR runs each thresholded at $p < 0.001$. For PPI effects of each ROI, the significant voxels were determined using intersection of the two TR runs and two methods each thresholded at $p < 0.01$. This slightly liberal threshold was chosen to ensure enough number of voxels survived in the conjunction of the four scenarios. The whole brain mask was determined as all voxels in the brain, including WM and CSF.

3. Results

3.1. Activations of the checkerboard task

Both TR runs showed highly significant activations in the visual cortex, as well as in the posterior portion of the thalamus (Figure 3A). The overlaps (Dice coefficients) of thresholded t maps between the two TR runs were as high as 0.7 (Figure 3B) at most of the shown t range percentile range. And Dice coefficients went down when only extremely activated voxels were thresholded. The visual cortex regions also showed high test-retest reliability (ICC greater than 0.7) (Figure 3C). However, the activations of the thalamus only showed small test-retest reliability around 0.2. The histograms of ICCs in the significant voxels and in the whole brain are shown on the right of Figure 3C.

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

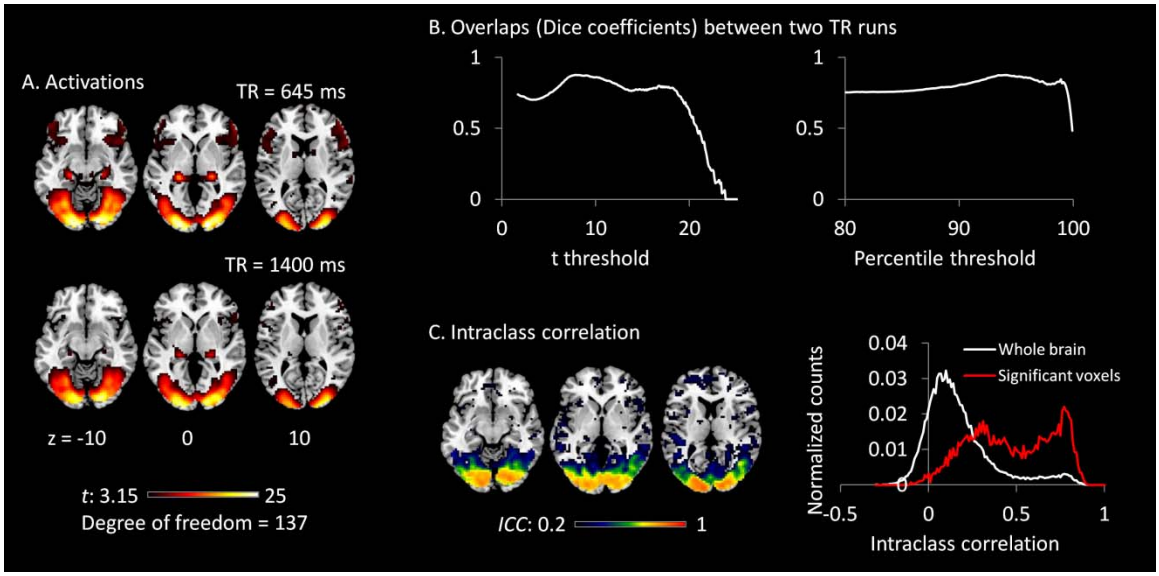

Figure 3 A) Activations (*t* maps) of visual checkerboard presentation for the 645 ms TR run (upper) and 1400 ms TR run (lower). The threshold t value corresponds to one-tailed significance at $p < 0.001$. B) Overlaps (Dice coefficients) between the two TR runs using t threshold (left) and percentile threshold (right). C) Test-retest reliability map (intraclass correlations, ICC) of activations between the two runs is shown on the left, which were thresholded at ICC > 0.2. The histograms of ICC of activations between the two TR runs in significant voxels and whole brain are shown on the right. The significant voxels were determined using intersection of the two runs each thresholded at $p < 0.001$.

3.2. Psychophysiological interactions

The voxel-wise PPI analysis of the LMOG and RMOG seeds conveyed very similar patterns. The PPI effects of the LMOG seed for the two TRs and two methods are shown in Figure 4. We first observed that even though spatial extents of PPI effects varied across the two TR runs and two PPI methods, the negative PPI effects in previously reported regions, i.e. supplementary motor area and higher visual cortex, could be observed from all four scenarios. The deconvolution method in 645 ms TR run had the smallest spatial extent and statistical significance, while the non-deconvolution method in 645 ms TR run had the largest spatial extent and strongest statistical significance. Both methods in TR of 1400 ms

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

showed similar spatial extent and significance levels. The last row in Figure 4 demonstrates the overlaps of negative effects in the four scenarios. Similar results were found in the analysis of the RMOG seed (Figure 5).

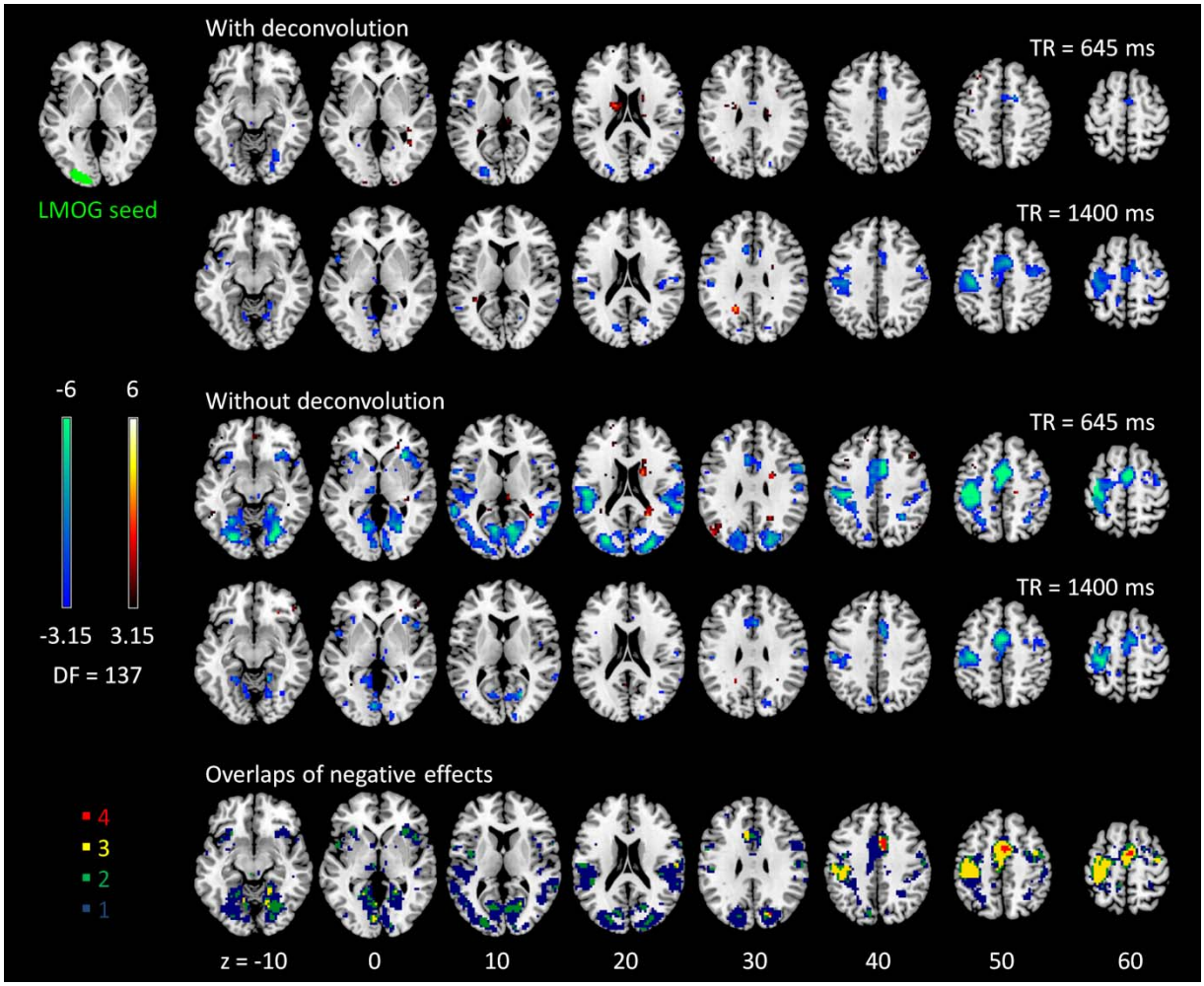

Figure 4 Psychophysiological interaction (PPI) results for the left middle occipital gyrus (LMOG) seed during checkerboard presentation in the two TR (repetition time) runs of 645 ms and TR 1400 ms. The resulting clusters were thresholded at $p < 0.001$ (approximated $t = 3.15$), with DF (degree of freedom) of 137. The last row illustrates the overlaps of the negative PPI results. Numbers on the bottom represent z coordinates in MNI (Montreal Neurology Institute) space.

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

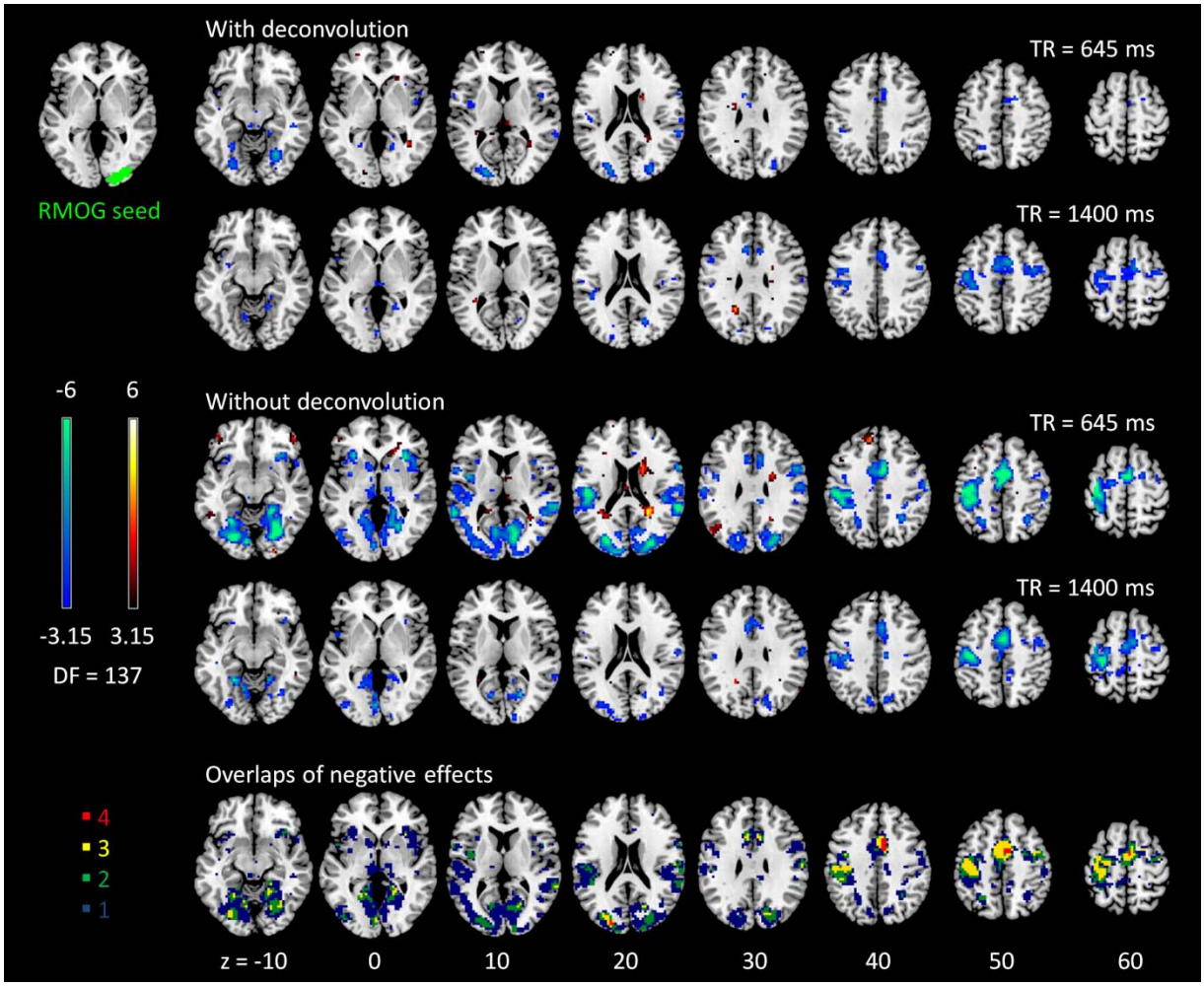

Figure 5 Psychophysiological interaction (PPI) results for the right middle occipital gyrus (RMOG) seed during checkerboard presentation in the two TR (repetition time) runs of 645 ms and TR 1400 ms. The resulting clusters were thresholded at $p < 0.001$ (approximated $t = 3.15$), with DF (degree of freedom) of 137. The last row illustrates the overlaps of the negative PPI results. Numbers on the bottom represent z coordinates in MNI (Montreal Neurology Institute) space.

The voxel-wise PPI analysis of the thalamus seed only showed significant effects in the 645 TR run, but with different brain regions with opposite effects (Figure 6). With deconvolution method, the thalamus seed showed significant positive PPI effects with middle cingulate gyrus, anterior portion of the thalamus, bilateral anterior insula, basal ganglia, and right fursiform gyrus. Whereas with non-

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

deconvolution method, the thalamus seed showed significant negative PPI effects with the bilateral occipital pole regions. There were no consistent results between two TR runs and two methods. Therefore subsequent analysis was only performed on the LMOG and RMOG seeds.

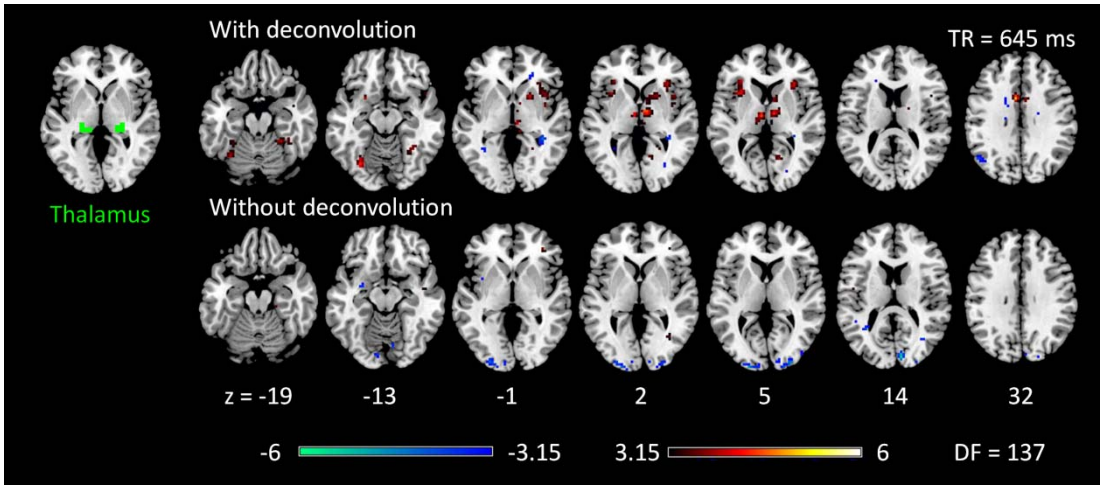

Figure 6 Psychophysiological interaction (PPI) results for the thalamus seed during checkerboard presentation in the TR (repetition time) run of 645 ms. There is no significant PPI effects of the thalamus seed in TR run of 1400 ms. The resulting clusters were thresholded at $p < 0.001$ (approximated $t = 3.15$), with DF (degree of freedom) of 137. Numbers on the bottom represent z coordinates in MNI (Montreal Neurology Institute) space.

We next performed ROI-based PPI analysis among 20 regions of visual thalamus and cortical visual areas (Figure 7). The 645 ms TR run showed more significant PPI effects than the 1400 ms TR run. And non-deconvolution method showed more significant PPI effects than the deconvolutio method. A prominent number of connectivity changes are between the bilateral FG1 regions and other lower level visual areas ranging from OC1, OC2, to OC4. We performed a conjunction analysis of PPI results across the four scenarios, and identified five connections with reduced connectivity in checkerboard than in fixation. The regions and connections are highlighted in Figure 8.

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

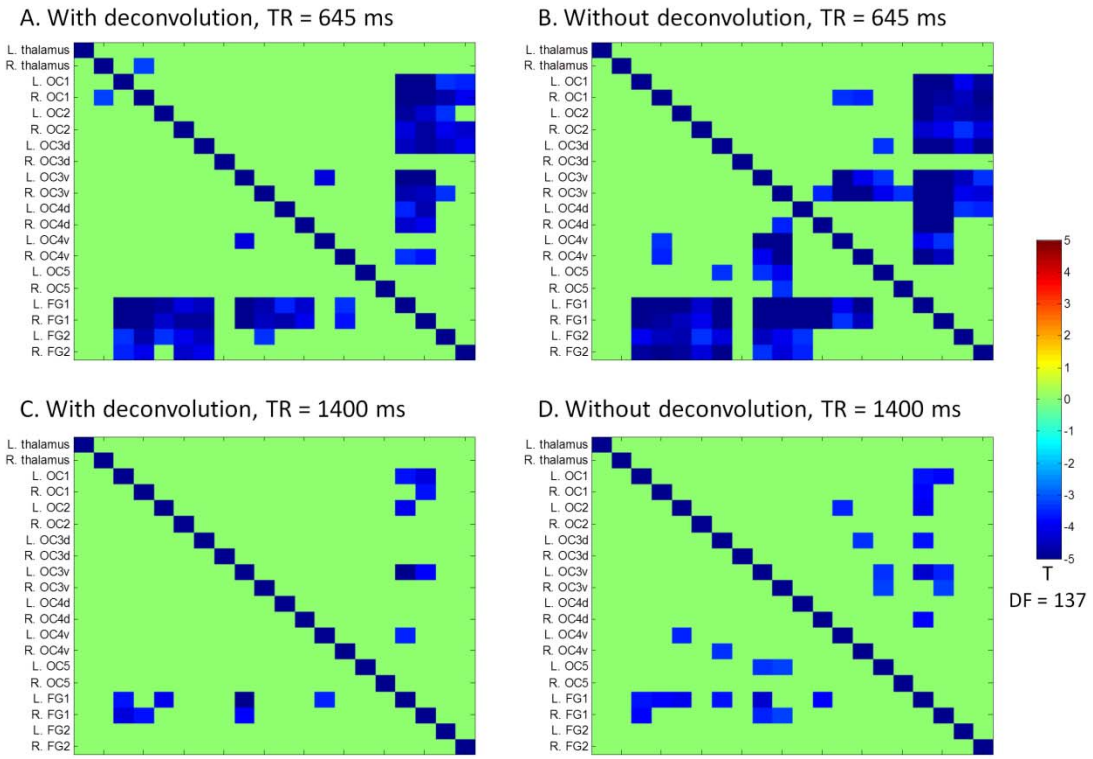

Figure 7 Matrices of psychophysiological interaction (PPI) results among the 20 regions of interest of visual thalamus and visual cortex for the two TR (repetition time) runs and two methods. The resulting clusters were thresholded at $p < 0.001$.

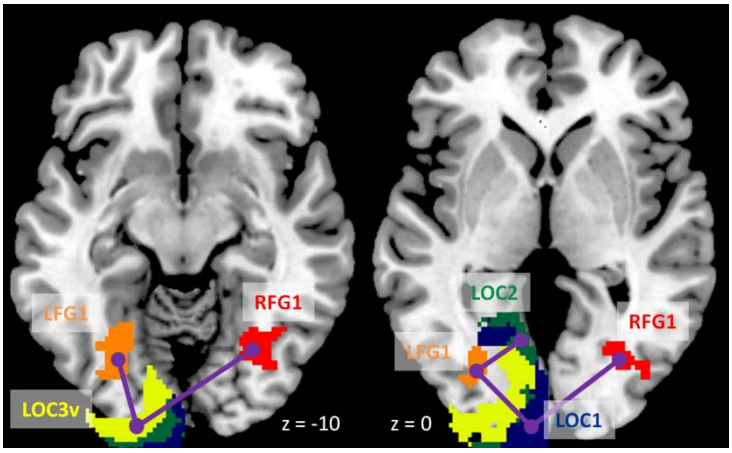

Figure 8 Illustration of consistently reduced connectivity during checkerboard presentation compared with fixation in the ROI-based (region of interest) psychophysiological interaction (PPI) analysis in the

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

two TR (repetition time) runs and two methods. Numbers on the bottom represent z coordinates in MNI (Montreal Neurology Institute) space.

3.3. Reproducibility of PPI effects

Since we observed similarities of spatial clusters and connectivity between the two TR runs, we next examined reproducibility of PPI effects by calculating Dice coefficients of thresholded statistical maps or PPI matrices between the two TR runs (Figure 9). For voxel-wise analysis of both LMOG and RMOG seeds, when varying t threshold, the non-deconvolution method showed higher level overlap compared with the deconvolution method (Figure 9A). When thresholding statistical maps with matched number of surviving voxels, a similar pattern could still be observed that the non-deconvolution method produced larger overlaps than the deconvolution method (Figure 9B). For the ROI-wise analysis, however, Dice coefficients were at similar level between two PPI methods at most t and percentile thresholds. But at very high t threshold or percentile thresholds, the deconvolution method seemed to produce larger overlaps (higher Dice coefficients) (Figure 9C and 9D).

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

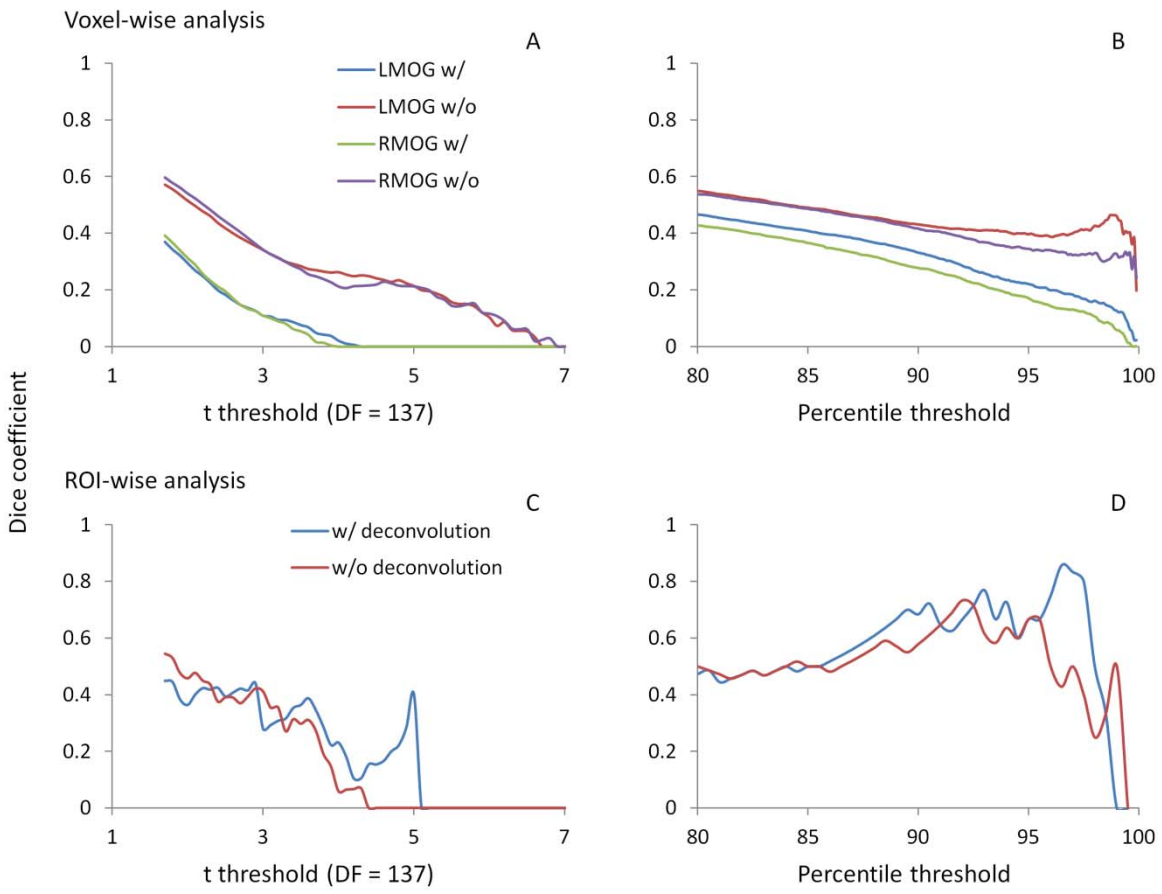

Figure 9 Dice coefficients of thresholded negative PPI effects between the two TR runs as functions of t threshold (A) and percentile threshold (B) for the two seeds and two PPI methods. The lowest t used for calculating overlap is 1.7, which approximately corresponds to $p < 0.05$. The largest percentile is 80 to 99.9 percentile, which is approximately corresponds to the largest proportions of voxels at $p < 0.05$.

3.4. Reliability of PPI effects

Lastly, we calculated ICC between the two TR runs to reflect reliability of PPI effects. The voxel-wise maps of ICC showed that there were typically low reliability in both methods and ROIs, even in the regions that showed consistent negative PPI effects (Supplementary Figure S1). We then plotted the histograms of ICCs in voxels from the whole brain (gray lines) and within regions that showed significant PPI effects (red lines) (Figure 10A through 10D). It turns out that the distributions of ICCs within

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

significant regions are only slightly different from the distributions of correlations in the whole brain, with means around 0.07. The distributions of ICCs were not different between deconvolution and non-deconvolution methods. Similar distributions of ICCs were also found for the ROI-wise analysis (Figure 10E, 10F, and supplementary Figure S2). We found five PPI effects that were consistently significant in both TR runs and methods. And the ICCs for the five effects were also small and close to zero.

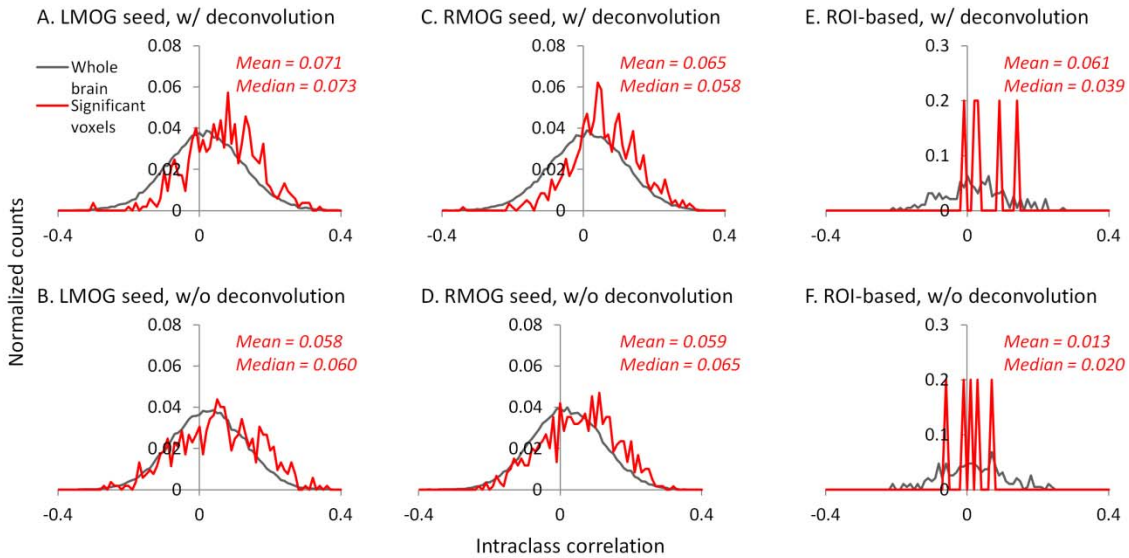

Figure 10 Histograms (normalized) of intraclass correlations of PPI effects between the two TR runs across the whole brain (gray lines) and in statistically significant voxels (red lines). The significant voxels were determined using intersection of the two runs and two methods each thresholded at $p < 0.01$. Left and right masks were calculated separately.

4. Discussion

By analyzing two separate runs of visual checkerboard task from a large sample (n = 138), the current study first replicated previously reported negative PPI effects between visual cortex and widespread brain regions, and then showed negative PPI effects among visual areas centered in the bilateral fusiform gyrus. By comparing results from two separate runs, we showed that group mean effects were largely

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reproducible; however, the inter-subject reliabilities of the PPI effects were typically low. By comparing the deconvolution and non-deconvolution PPI methods, we demonstrated that the results by the two methods were in general very similar, but the non-deconvolution produced larger statistical effects and spatial extent. The non-deconvolutionmethod may also increase results overlaps between the two runs in some circumstances compared with the deconvolution method.

4.1. Functional connectivity during checkerboard stimulation

The voxel-wise analysis of the LMOG and RMOG seeds replicated our previous results which only analyzed a sub-set of 26 subjects (Di et al., 2017b, 2015). In our previous work (Di et al., 2017b) we could only identify significant PPI effects using the RMOG seed, while the current study demonstrated similar PPI effects using both LMOG and RMOG seeds. Furthermore, we illustrated that the spatial extent of regions that showed reduced connectivity with the MOG seed could be much larger and extended to other brain regions such as the insula and bilateral sensorimotor cortex. This further suggests a more functional segregation between the visual cortex and other brain systems during such a simple visual stimulation task compared with fixation. The current study also extended previous study by analyzing task modulated connectivity effects among cytoarchitectonically defined visual areas. Reduced functional connectivity was also observed among many visual areas, with the bilateral FG1 as hub regions. FG1 is the most posterior portion of the fusiform gyrus, which just laid anterior to the occipital cortex (Caspers et al., 2013), It was thought a transition zone between lower retinotopic visual areas and higher category specific brain areas, which integrate information from different retinotopic visual areas to higher category specific brain areas (Caspers et al., 2014). Therefore, it is reasonable to see that the FG1 showed reduced functional connectivity with many lower visual areas in the checkerboard condition, because the simple stimuli cannot form a meaningful percept of a specific category.

The thalamus is a critical subcortical structure in the brain, which not only relay sensory information to the cortex, but also thought to mediate corticocortical communications (Guillery and Sherman, 2002; Saalmann and Kastner, 2011). The PPI analysis of the thalamus, however, did not show

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

consistent effects in different TR runs, and different methods. The reason may be that the visual thalamus is small in size compared with cortical visual areas, and the signals in the thalamus are not reliable enough. The current results do suggest some reduced connectivity between the visual thalamus to the primary visual cortex, and increased connectivity between the visual thalamus to the anterior portion of the thalamus, basal ganglia, and insula. However, the results are weak and unreliable, especially considering that the current analysis had included 138 subjects.

4.2. Reproducibility and reliability of PPI effects

To our knowledge, the current study is the first one to evaluate reproducibility and reliability on PPI effects. The current analysis did not only reproduce the results reported previously (Di et al., 2017b), but also enabled us to examine reproducibility between two runs. Although the two runs were scanned using different parameters, most importantly the temporal and spatial resolutions, the patterns of PPI effects turned out to be quite similar between the two runs. The run with 645 ms TR seemed to generate larger spatial extent in voxel-wise analysis and more significant results in ROI-wise analysis. This is consistent with our prediction, because there are more time points in the 645 ms TR run than in the 1400 ms TR run, which could yield higher statistical power. We do notice that in some scenarios, i.e. voxel-wise analysis with deconvolution, the PPI results in 645 ms TR run had smaller effect size and spatial extent. This might be due to failure of deconvolution, which will be discussed below.

On the other hand, the results also indicated that inter-subject reliabilities are typically low (around 0.07) no matter which PPI method was used. The low reliability could not be simply due to different scanning parameters in the two runs, because simple task activations still showed reasonably high reliability. In addition, the reliability of PPI effects in the current analysis are also much lower than previous test-retest studies on task activations (Plichta et al., 2012; Raemaekers et al., 2007) and on resting-state functional connectivity (Guo et al., 2012; Zuo et al., 2010b). It is reasonable, because PPI measures higher order interaction effect between main effects of task activations and simple (task-free) functional connectivity, so that PPI results should be less reliable than the main effects of activations and

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

connectivity. The short scan length may also explain the low reliability. There are in total 60 s (20 x 3) for the checkerboard presentation condition, which is much less than a typical resting-state scan (usually 5 min or longer). However, it is a typical experiment length for a task fMRI study, especially for a strong effect such as flickering checkerboard. Here, we are not overgeneralizing our results by suggesting that the reliability of other PPI studies are only as low as 0.07. However, one needs to consider the fact of lower reliability in PPI results than those in activation results when designing a task when the goal of a study is to examine task related connectivity. And practically for a task fMRI experiment, comparisons are usually made between two well-designed experimental conditions, rather than between an experimental condition and a fixation condition as used in the current analysis. In such a situation, the PPI effects would be much smaller, thus requiring much longer overall scan time.

4.3. Deconvolution and PPI

The PPI results using both the deconvolution and non-deconvolution methods are in general very similar. This is consistent with the simulation showing that the PPI term calculated from non-deconvolution method is very similar to the hypothetical PPI term with a known neural activity for a block-designed task. When comparing the differences of PPI results with these two methods, the non-deconvolution method seems to generate larger statistical effects and larger spatial extent or number of significant effects. The non-deconvolution method also increased the Dice coefficients of thresholded PPI maps between the two TR runs. However, the Dice coefficients of thresholded PPI matrices between the two TR runs are quite similar between the two PPI methods, and the deconvolution method may be even benefiting at higher thresholds. These results highlighted the uncertainty of deconvolution method in PPI analysis.

To gain further insight to the cases of deconvolution failure, we calculated correlations of PPI terms between deconvolution and non-deconvolution methods for the LMOG and RMOG seeds (Figure 11A). In both TR runs, the distributions of correlations centered approximately on 0.7, and there were outliers whose correlations were only 0.2 or 0.3. This is in contrast with the simulation results (Figure 1B, 40 sec cycle), where the correlations were around 0.9. The lower correlations of PPI terms from

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

empirical data compared with the simulations imply that there might be some uncountable variations introduced during the deconvolution/convolution of real fMRI data. Indeed, deconvolution is always a practical problem that to recover underlying signals from some recorded measures, rather than a simple mathematical problem as depicted in equation 2. In such a scenario, measurement noises need to be taken into account in the deconvolution model. In the context of fMRI, the goal of deconvolution is to recover neuronal activities from observed BOLD signals, where there are plenty of noises during MRI recording. The deconvolution should be expressed as follows with an additional error term:

$$x_{Physio} = z_{Physio} * hrf + \varepsilon \qquad (5)$$

In this circumstance, some noises would be removed during deconvolution so that a signal deconvolved and convolved back with a HRF will no longer be the same as the original signal. The noise characteristics and regularization methods for recovering z_{Physio} become critical to the success of deconvolution.

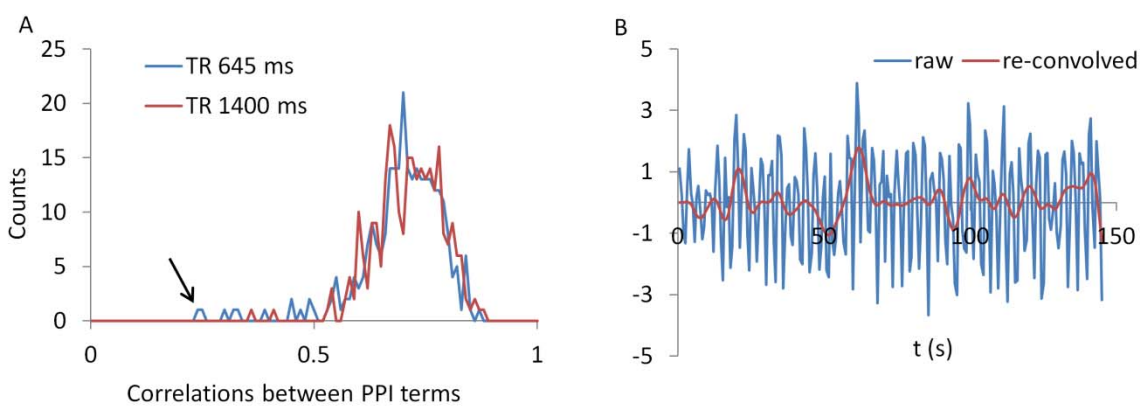

Figure 11 A) Histograms of correlations between PPI terms with and without deconvolution across all subjects from both the LMOG and RMOG ROIs for the two TR runs. B) For the worst case as spotted by the black arrow in A), we show the raw time series and the time series with deconvolution and re-convolution with hemodynamic response function.

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We identified the worst case in Figure 11A (black arrow indicated), and deconvolved and reconvovled it with HRF using SPM's method (Figure 11B). The raw and reconvovled signals look dramatically different, with the reconvolved signal resembling a smoothed version of the original signal. Smoothness is indeed the case for the SPM version of deconvolution (Gitelman et al., 2003), because it utilizes regularization to suppress high frequency components of cosine basis functions those were used to approximate z_{physio}. To directly illustrate the point, we performed fast Fourier transformation on the time series of the RMOG for all the subjects on the raw, deconvolved, and reconvolved time series for the two TR runs (Figure 12). It could be seen that after deconvolution, high frequency components have been suppressed in both TR runs. We note that there is a black line which shows higher power between frequency of 0.2 to 0.4 Hz in the raw data plot of TR = 645 ms run, which coincides to be the outlier observed in Figure 11. The high frequency component was suppressed, so that the reconvolved signal looks smooth.

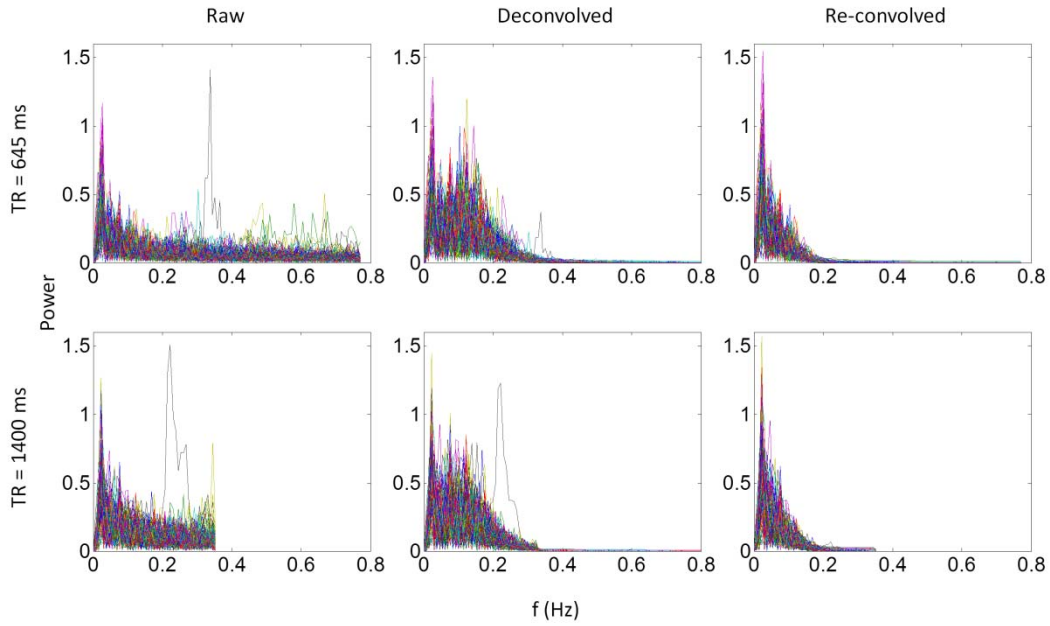

Figure 12 Power spectrums of time series from the right middle occipital gyrus seed for each of the 138 subjects for the 645 ms run (upper panels) and 1400 ms run (lower panels). Each line in a plot represents

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

one subject. Left, middle, and right panels show the power spectrum of the raw, deconvolved, and re-convolved time series, respectively.

SPM's deconvolution method explicitly suppresses high frequency components with the intention that the hemodynamic response is slow therefore high frequency components may represent noises. But this may overly smooth the data, and remove useful information in higher frequency bands, thus making PPI results with the deconvolution method less sensitive than those with the direct PPI method. This problem may be more severe for short TR data, because there are more high frequency components in the data. On the other hand, high frequency signals in BOLD have been increasingly recognized as functionally meaningful (Chen and Glover, 2015; Gohel and Biswal, 2015; Lewis et al., 2016), and high frequency components may be critical for connectivity dynamics. Given that multiband imaging technique has made fMRI sampling rate much faster, proper treatment of high frequency signals may be critical in deconvolution of fMRI signals and connectivity analysis in general.

Given the facts that the two PPI methods can generate similar results for the current block-designed task and the non-deconvolution method may increase statistical power, we lean toward a conclusion that the non-deconvolution PPI method may be a better choice for a block-designed task. This is in line with the recommendation by FSL (O'Reilly et al., 2012). Of course, deconvolution is still necessary for an event-related task design, because the PPI terms calculated from the non-deconvolution method are dramatically different from those calculated from the deconvolution method (Figure 1). But it's worth mentioning that it has been suggested that the beta series method (Rissman et al., 2004) might be a better alternative for event-related designed data (Cisler et al., 2014). Lastly, there are indeed many variety of deconvolution methods (Havlicek et al., 2011; Makni et al., 2008; Wu et al., 2013), and some of the methods may be more suitable for fMRI signals and PPI analysis. But systematic analyses comparing these different methods are needed in the future.

4.4. Practical implications on PPI analysis

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Task related functional connectivity as measured by PPI analysis is typically much smaller, in terms of effect size, reproducibility, and reliability, than simple task activations. To ensure enough statistical power and reliability, a larger sample size than typical activation studies and enough scan length for each subject are required. The design for an fMRI task needs to consider scan length as a critical factor, if the goal of the study is to examine task related connectivity. To date, it is still largely unknown how long the scan is needed for reliability capture task related connectivity. We can only get some insight from resting-state connectivity research, where large scale test-retest datasets are available (Biswal et al., 2010; Zuo et al., 2014). In resting-state literature, it has been suggested that at least five minutes of scan is needed for reliability estimate functional connectivity (Birn et al., 2013; Van Dijk et al., 2010). Then at least five minutes of scan length for a single task condition is needed for task based fMRI. If the PPI effects are going to be compared between two experimental conditions, which is usually the case for a well-designed cognitive neuroimaging study, the required scan length would be much longer. Of course, direct examinations of the effect of scan length on task related connectivity estimates are still needed in future research.

Secondly, the PPI method tasks advantages of the dynamic aspect of the BOLD signals. Therefore, it's preferable to adopt faster sampling rate to capture temporal dynamics, which may in turn lead to sacrifice of other aspects of signals, e.g. spatial resolution. The current results support the idea that shorter TR may be beneficial for PPI analysis. Of course, faster sampling rate could be accomplished by new developments of MRI techniques such as multi-band acquisition (Feinberg and Yacoub, 2012). However, the current results also suggested some pitfalls of using short TR data. The currently used HRF models and deconvolution method may be not quite suitable for fast TR data, so that the PPI method with deconvolution may fail in some cases in short TR data. More work is still needed to validate and optimize models on high speed fMRI data.

5. Conclusion

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We demonstrated that the deconvolution and non-deconvolution PPI methods generated similar results on a simple block-designed task. The deconvolution method may be beneficial in terms of statistical power and reproducibility. Taken together, deconvolution may be not necessary for PPI analysis on block-designed fMRI data. When using a large sample, group mean PPI effects are reproducible; however, inter-subject reliabilities of the PPI effects are quite limited. Systematic evaluations on scan length and reliability may be necessary before studying inter-subject differences or group differences of PPI effects.

Acknowledgement

This research was supported by National Institute of Health grants R01 AG032088 and R01 DA038895.

Reference

Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., Zilles, K., 2000. Brodmann's Areas 17 and 18 Brought into Stereotaxic Space—Where and How Variable? Neuroimage 11, 66–84. doi:10.1006/nimg.1999.0516

Birn, R.M., Molloy, E.K., Patriat, R., Parker, T., Meier, T.B., Kirk, G.R., Nair, V.A., Meyerand, M.E., Prabhakaran, V., 2013. The effect of scan length on the reliability of resting-state fMRI connectivity estimates. Neuroimage 83, 550–558. doi:10.1016/j.neuroimage.2013.05.099

Biswal, B.B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski, A.-M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S., Kiviniemi, V.J., Kötter, R., Li, S.-J., Lin, C.-P., Lowe, M.J., Mackay, C., Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H.S., McMahon, K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Petersen, S.E., Riedl, V., Rombouts, S.A.R.B., Rypma, B., Schlaggar, B.L., Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.-J., Veijola, J., Villringer,

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A., Walter, M., Wang, L., Weng, X.-C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.-F., Zhang, H.-Y., Castellanos, F.X., Milham, M.P., 2010. Toward discovery science of human brain function. Proc. Natl. Acad. Sci. U. S. A. 107, 4734–9. doi:10.1073/pnas.0911855107

Caspers, J., Zilles, K., Amunts, K., Laird, A.R., Fox, P.T., Eickhoff, S.B., 2014. Functional characterization and differential coactivation patterns of two cytoarchitectonic visual areas on the human posterior fusiform gyrus. Hum. Brain Mapp. 35, 2754–2767. doi:10.1002/hbm.22364

Caspers, J., Zilles, K., Eickhoff, S.B., Schleicher, A., Mohlberg, H., Amunts, K., 2013. Cytoarchitectonical analysis and probabilistic mapping of two extrastriate areas of the human posterior fusiform gyrus. Brain Struct. Funct. 218, 511–526. doi:10.1007/s00429-012-0411-8

Chen, J.E., Glover, G.H., 2015. BOLD fractional contribution to resting-state functional connectivity above 0.1Hz. Neuroimage 107, 207–218. doi:10.1016/j.neuroimage.2014.12.012

Cisler, J.M., Bush, K., Steele, J.S., 2014. A comparison of statistical methods for detecting context-modulated functional connectivity in fMRI. Neuroimage 84, 1042–1052. doi:10.1016/j.neuroimage.2013.09.018

Di, X., Biswal, B.B., 2015. Characterizations of resting-state modulatory interactions in the human brain. J. Neurophysiol. 114, 2785–96. doi:10.1152/jn.00893.2014

Di, X., Fu, Z., Chan, S.C., Hung, Y.S., Biswal, B.B., Zhang, Z., 2015. Task-related functional connectivity dynamics in a block-designed visual experiment. Front. Hum. Neurosci. 9, 1–11. doi:10.3389/fnhum.2015.00543

Di, X., Huang, J., Biswal, B.B., 2017a. Task modulated brain connectivity of the amygdala: a meta-analysis of psychophysiological interactions. Brain Struct. Funct. 222, 619–634. doi:10.1007/s00429-016-1239-4

Di, X., Reynolds, R.C., Biswal, B.B., 2017b. Imperfect (de)convolution may introduce spurious psychophysiological interactions and how to avoid it. Hum. Brain Mapp. 38, 1723–1740. doi:10.1002/hbm.23413

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dubois, J., Adolphs, R., 2016. Building a Science of Individual Differences from fMRI. Trends Cogn. Sci. 20, 425–443. doi:10.1016/j.tics.2016.03.014

Feinberg, D.A., Yacoub, E., 2012. The rapid development of high speed, resolution and precision in fMRI. Neuroimage 62, 720–725. doi:10.1016/j.neuroimage.2012.01.049

Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6, 218–29.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302. doi:10.1016/S1053-8119(03)00202-7

Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., Turner, R., 1996. Movement-related effects in fMRI time-series. Magn. Reson. Med. 35, 346–55. doi:DOI 10.1002/mrm.1910350312

Gitelman, D.R., Penny, W.D., Ashburner, J., Friston, K.J., 2003. Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. Neuroimage 19, 200–7.

Gohel, S.R., Biswal, B.B., 2015. Functional Integration Between Brain Regions at Rest Occurs in Multiple-Frequency Bands. Brain Connect. 5, 23–34. doi:10.1089/brain.2013.0210

Guillery, R.W., Sherman, S.M., 2002. Thalamic Relay Functions and Their Role in Corticocortical Communication. Neuron 33, 163–175.

Guo, C.C., Kurth, F., Zhou, J., Mayer, E.A., Eickhoff, S.B., Kramer, J.H., Seeley, W.W., 2012. One-year test–retest reliability of intrinsic connectivity network fMRI in older adults. Neuroimage 61, 1471–1483. doi:10.1016/j.neuroimage.2012.03.027

Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21, 1639–1651. doi:10.1016/j.neuroimage.2003.11.029

Havlicek, M., Friston, K.J., Jan, J., Brazdil, M., Calhoun, V.D., 2011. Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering. Neuroimage 56, 2109–2128.

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

doi:10.1016/j.neuroimage.2011.03.005

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage. doi:10.1016/j.neuroimage.2011.09.015

Kujovic, M., Zilles, K., Malikovic, A., Schleicher, A., Mohlberg, H., Rottschy, C., Eickhoff, S.B., Amunts, K., 2013. Cytoarchitectonic mapping of the human dorsal extrastriate cortex. Brain Struct. Funct. 218, 157–172. doi:10.1007/s00429-012-0390-9

Lewis, L.D., Setsompop, K., Rosen, B.R., Polimeni, J.R., 2016. Fast fMRI can detect oscillatory neural activity in humans. Proc. Natl. Acad. Sci. 201608117. doi:10.1073/pnas.1608117113

Makni, S., Beckmann, C., Smith, S., Woolrich, M., 2008. Bayesian deconvolution fMRI data using bilinear dynamical systems. Neuroimage 42, 1381–1396. doi:10.1016/j.neuroimage.2008.05.052

Malikovic, A., Amunts, K., Schleicher, A., Mohlberg, H., Eickhoff, S.B., Wilms, M., Palomero-Gallagher, N., Armstrong, E., Zilles, K., 2006. Cytoarchitectonic Analysis of the Human Extrastriate Cortex in the Region of V5/MT+: A Probabilistic, Stereotaxic Map of Area hOc5. Cereb. Cortex 17, 562–574. doi:10.1093/cercor/bhj181

Nooner, K.B., Colcombe, S.J., Tobe, R.H., Mennes, M., Benedict, M.M., Moreno, A.L., Panek, L.J., Brown, S., Zavitz, S.T., Li, Q., Sikka, S., Gutman, D., Bangaru, S., Schlachter, R.T., Kamiel, S.M., Anwar, A.R., Hinz, C.M., Kaplan, M.S., Rachlin, A.B., Adelsberg, S., Cheung, B., Khanuja, R., Yan, C., Craddock, C.C., Calhoun, V., Courtney, W., King, M., Wood, D., Cox, C.L., Kelly, A.M.C., Di Martino, A., Petkova, E., Reiss, P.T., Duan, N., Thomsen, D., Biswal, B., Coffey, B., Hoptman, M.J., Javitt, D.C., Pomara, N., Sidtis, J.J., Koplewicz, H.S., Castellanos, F.X., Leventhal, B.L., Milham, M.P., 2012. The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. Front. Neurosci. 6, 152. doi:10.3389/fnins.2012.00152

O'Reilly, J.X., Woolrich, M.W., Behrens, T.E.J., Smith, S.M., Johansen-Berg, H., 2012. Tools of the Trade: Psychophysiological Interactions and Functional Connectivity. Soc. Cogn. Affect. Neurosci. nss055-. doi:10.1093/scan/nss055

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Plichta, M.M., Schwarz, A.J., Grimm, O., Morgen, K., Mier, D., Haddad, L., Gerdes, A.B.M., Sauer, C., Tost, H., Esslinger, C., Colman, P., Wilson, F., Kirsch, P., Meyer-Lindenberg, A., 2012. Test–retest reliability of evoked BOLD signals from a cognitive–emotive fMRI test battery. Neuroimage 60, 1746–1758. doi:10.1016/j.neuroimage.2012.01.129

Raemaekers, M., Vink, M., Zandbelt, B., van Wezel, R.J.A., Kahn, R.S., Ramsey, N.F., 2007. Test–retest reliability of fMRI activation during prosaccades and antisaccades. Neuroimage 36, 532–542. doi:10.1016/j.neuroimage.2007.03.061

Rissman, J., Gazzaley, A., D'Esposito, M., 2004. Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage 23, 752–63. doi:10.1016/j.neuroimage.2004.06.035

Roebroeck, A., Formisano, E., Goebel, R., 2011. The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. Neuroimage 58, 296–302. doi:10.1016/j.neuroimage.2009.09.036

Rombouts, S. a, Barkhof, F., Hoogenraad, F.G., Sprenger, M., Scheltens, P., 1998. Within-subject reproducibility of visual activation patterns with functional magnetic resonance imaging using multislice echo planar imaging. Magn. Reson. Imaging 16, 105–113.

Rottschy, C., Eickhoff, S.B., Schleicher, A., Mohlberg, H., Kujovic, M., Zilles, K., Amunts, K., 2007. Ventral visual cortex in humans: Cytoarchitectonic mapping of two extrastriate areas. Hum. Brain Mapp. 28, 1045–1059. doi:10.1002/hbm.20348

Saalmann, Y.B., Kastner, S., 2011. Cognitive and Perceptual Functions of the Visual Thalamus. Neuron 71, 209–223. doi:10.1016/j.neuron.2011.06.027

Taylor, P.A., Gohel, S., Di, X., Walter, M., Biswal, B.B., 2012. Functional covariance networks: obtaining resting-state networks from intersubject variability. Brain Connect. 2, 203–17. doi:10.1089/brain.2012.0095

Van Dijk, K.R.A., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., Buckner, R.L., 2010. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

optimization. J. Neurophysiol. 103, 297–321. doi:10.1152/jn.00783.2009

Vul, E., Harris, C., Winkielman, P., Pashler, H., 2009. Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. Perspect. Psychol. Sci. 4, 274–290. doi:10.1111/j.1745-6924.2009.01125.x

Wu, G.-R., Liao, W., Stramaglia, S., Ding, J.-R., Chen, H., Marinazzo, D., 2013. A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data. Med. Image Anal. 17, 365–74. doi:10.1016/j.media.2013.01.003

Yuan, R., Di, X., Taylor, P.A., Gohel, S., Tsai, Y.-H., Biswal, B.B., 2016. Functional topography of the thalamocortical system in human. Brain Struct. Funct. 221, 1971–1984. doi:10.1007/s00429-015-1018-7

Zuo, X.-N., Anderson, J.S., Bellec, P., Birn, R.M., Biswal, B.B., Blautzik, J., Breitner, J.C.., Buckner, R.L., Calhoun, V.D., Castellanos, F.X., Chen, A., Chen, B., Chen, J., Chen, X., Colcombe, S.J., Courtney, W., Craddock, R.C., Di Martino, A., Dong, H.-M., Fu, X., Gong, Q., Gorgolewski, K.J., Han, Y., He, Y., He, Y., Ho, E., Holmes, A., Hou, X.-H., Huckins, J., Jiang, T., Jiang, Y., Kelley, W., Kelly, C., King, M., LaConte, S.M., Lainhart, J.E., Lei, X., Li, H.-J., Li, K., Li, K., Lin, Q., Liu, D., Liu, J., Liu, X., Liu, Y., Lu, G., Lu, J., Luna, B., Luo, J., Lurie, D., Mao, Y., Margulies, D.S., Mayer, A.R., Meindl, T., Meyerand, M.E., Nan, W., Nielsen, J.A., O'Connor, D., Paulsen, D., Prabhakaran, V., Qi, Z., Qiu, J., Shao, C., Shehzad, Z., Tang, W., Villringer, A., Wang, H., Wang, K., Wei, D., Wei, G.-X., Weng, X.-C., Wu, X., Xu, T., Yang, N., Yang, Z., Zang, Y.-F., Zhang, L., Zhang, Q., Zhang, Z., Zhang, Z., Zhao, K., Zhen, Z., Zhou, Y., Zhu, X.-T., Milham, M.P., 2014. An open science resource for establishing reliability and reproducibility in functional connectomics. Sci. Data 1, 140049. doi:10.1038/sdata.2014.49

Zuo, X.-N., Di Martino, A., Kelly, C., Shehzad, Z.E., Gee, D.G., Klein, D.F., Castellanos, F.X., Biswal, B.B., Milham, M.P., 2010a. The oscillating brain: Complex and reliable. Neuroimage 49, 1432–1445. doi:10.1016/j.neuroimage.2009.09.037

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Zuo, X.-N., Kelly, C., Adelstein, J.S., Klein, D.F., Castellanos, F.X., Milham, M.P., 2010b. Reliable intrinsic connectivity networks: Test–retest evaluation using ICA and dual regression approach. Neuroimage 49, 2163–2177. doi:10.1016/j.neuroimage.2009.10.080

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure legend

Figure 1 Simulations of the correlations between PPI terms calculated from deconvolution and non-deconvolution methods. Panel A illustrates different task designs that were used for the simulation. Each column represents a task design. 0 of x axis represents the event-related design, with 1 time bin (2 s) of trial condition and 5 time bins (10 s) of baseline condition. The remaining columns show block designs with different frequencies of repetition. For example, 80 secs cycle means 40-s on and 40-s off of the task condition related to baseline. A physiological variable was generated using Gaussian random noise for 1,000 times. Penal B shows boxplots of correlations between PPI terms calculated from deconvolution and non-deconvolution methods with the 1,000 simulations. The code for this simulation can be found at: https://github.com/dixy0/PPI_correlation_demo .

Figure 2 Examples of PPI terms calculated by deconvolution and non-deconvolution methods for the two TR runs.

Figure 3 A) Activations (*t* maps) of visual checkerboard presentation for the 645 ms TR run (upper) and 1400 ms TR run (lower). The threshold t value corresponds to one-tailed significance at $p < 0.001$. B) Overlaps (Dice coefficients) between the two TR runs using t threshold (left) and percentile threshold (right). C) Test-retest reliability map (intraclass correlations, ICC) of activations between the two runs is shown on the left, which were thresholded at ICC > 0.2. The histograms of ICC of activations between the two TR runs in significant voxels and whole brain are shown on the right. The significant voxels were determined using intersection of the two runs each thresholded at $p < 0.001$.

Figure 4 Psychophysiological interaction (PPI) results for the left middle occipital gyrus (LMOG) seed during checkerboard presentation in the two TR (repetition time) runs of 645 ms and TR 1400 ms. The resulting clusters were thresholded at $p < 0.001$ (approximated $t = 3.15$), with DF (degree of freedom) of

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

137. The last row illustrates the overlaps of the negative PPI results. Numbers on the bottom represent z coordinates in MNI (Montreal Neurology Institute) space.

Figure 5 Psychophysiological interaction (PPI) results for the right middle occipital gyrus (RMOG) seed during checkerboard presentation in the two TR (repetition time) runs of 645 ms and TR 1400 ms. The resulting clusters were thresholded at $p < 0.001$ (approximated $t = 3.15$), with DF (degree of freedom) of 137. The last row illustrates the overlaps of the negative PPI results. Numbers on the bottom represent z coordinates in MNI (Montreal Neurology Institute) space.

Figure 6 Psychophysiological interaction (PPI) results for the thalamus seed during checkerboard presentation in the TR (repetition time) run of 645 ms. There is no significant PPI effects of the thalamus seed in TR run of 1400 ms. The resulting clusters were thresholded at $p < 0.001$ (approximated $t = 3.15$), with DF (degree of freedom) of 137. Numbers on the bottom represent z coordinates in MNI (Montreal Neurology Institute) space.

Figure 7 Matrices of psychophysiological interaction (PPI) results among the 20 regions of interest of visual thalamus and visual cortex for the two TR (repetition time) runs and two methods. The resulting clusters were thresholded at $p < 0.001$.

Figure 8 Illustration of consistently reduced connectivity during checkerboard presentation compared with fixation in the ROI-based (region of interest) psychophysiological interaction (PPI) analysis in the two TR (repetition time) runs and two methods. Numbers on the bottom represent z coordinates in MNI (Montreal Neurology Institute) space.

bioRxiv preprint doi: https://doi.org/10.1101/133454; this version posted June 7, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 9 Dice coefficients of thresholded negative PPI effects between the two TR runs as functions of t threshold (A) and percentile threshold (B) for the two seeds and two PPI methods. The lowest t used for calculating overlap is 1.7, which approximately corresponds to $p < 0.05$. The largest percentile is 80 to 99.9 percentile, which is approximately corresponds to the largest proportions of voxels at $p < 0.05$.

Figure 10 Histograms (normalized) of intraclass correlations of PPI effects between the two TR runs across the whole brain (gray lines) and in statistically significant voxels (red lines). The significant voxels were determined using intersection of the two runs and two methods each thresholded at $p < 0.01$. Left and right masks were calculated separately.

Figure 11 A) Histograms of correlations between PPI terms with and without deconvolution across all subjects from both the LMOG and RMOG ROIs for the two TR runs. B) For the worst case as spotted by the black arrow in A), we show the raw time series and the time series with deconvolution and re-convolution with hemodynamic response function.

Figure 12 Power spectrums of time series from the right middle occipital gyrus seed for each of the 138 subjects for the 645 ms run (upper panels) and 1400 ms run (lower panels). Each line in a plot represents one subject. Left, middle, and right panels show the power spectrum of the raw, deconvolved, and re-convolved time series, respectively.