The Unreliable Influence of Noise Normalization on the Reliability of Neural Dissimilarity

J. Brendan Ritchie (brendan.ritchie@kuleuven.be)

Haemy Lee Masson (haemy.leemasson@kuleuven.be)

Stefania Bracci (stefania.bracci@kuleuven.be)

Hans Op de Beeck (hans.opdebeeck@kuleuven.be)
Laboratory of Biological Psychology, Brain and Cognition, KU Leuven
102 Tiensestraat – Box 3714, 3000 Leuven, Flemish Brabant, Belgium

Abstract:

Representational similarity analysis (RSA) is increasingly part of the standard analytic toolkit in neuroimaging. Core to RSA is the measuring of neural dissimilarity between the response patterns for different conditions to construct neural representational dissimilarity matrices (RDMs). It has been proposed that noise normalizing these patterns, and using cross-validated distances as a dissimilarity measure, is superior for characterizing the structure of neural RDMs. This assessment has been motivated by improvement in within-subject neural dissimilarity after noise normalization. However, between-subject reliability is more directly related to determining the amount of explainable variance, and the evaluation of observed effect sizes when they are correlated with behavioral or model RDMs. Across three datasets we did not find that noise normalization consistently boosts within-subject reliability, between-subject reliability or correlations with behavioral or model RDMs. Overall, our results provide equivocal support for the utility of noise normalization to RSA.

Keywords: representational similarity analysis; noise normalization; decoding; multivariate pattern analysis

Introduction

Representational similarity analysis (RSA) is increasingly part of the standard analytic toolkit utilized in neuroimaging. Central to RSA is the measuring of neural dissimilarity between the neural response patterns for different conditions to construct neural representational dissimilarity matrices (RDMs). Previously, it has been proposed that noise normalizing these patterns, and using cross-validated distances as a dissimilarity measure, is superior for characterizing the structure of neural RDMs (Walther et al., 2016). This assessment has been motivated by resulting suggesting improvement in within-subject neural dissimilarity after noise normalization.

However, when it comes to carrying out RSA, between-subject reliability is also related to determining the amount of explainable variance, and the evaluation of observed effect sizes when they are correlated with behavioral or model RDMs. Therefore, we sought to further investigate this issue by revisiting the impact of noise normalization on the reliability of measures of neural dissimilarity. Across three datasets, we did not find that noise normalization consistently boosted within-subject reliability, between-subject reliability or correlations with behavioral or model RDMs. Overall, our results provide equivocal support for the utility of noise normalization to RSA.

Methods

Datasets

Datasets from three previously published studies were reanalyzed. Preprocessing (e.g. motion correction) and first-level analysis of the data followed standard pipelines (e.g. SPM). Further methodological details can be found in those works. Notably, unlike in the original studies, in the present reanalysis data was not smoothed.

First, for Dataset 1 (D1), neural responses (N = 10) from primarily visual cortex (V1) were collected for 16 square-wave gratings varying in 4 levels of orientation and spatial frequency (Ritchie & Op de Beeck, 2019). Second, for Dataset 2 (D2), neural responses (N = 15) from object-selective lateral occipitotemporal cortex (object-LOTC) were collected for 54 natural images from 6 categories and 9 shape types (Bracci & Op de Beeck, 2016). Finally, for Dataset 3 (D3), neural responses (N = 21) from functionally responsive temporal parietal junction (TPJ) were collected for 75 videos of actors engaging in either social (39) or non-social (36) interactions with confederates or objects, respectively (Lee Masson et al. 2018).

Dissimilarity Measures

Four measures were used to calculate pairwise dissimilarity (Walther et al., 2016). First, the **Correlation Distance (COR)** dissimilarity was estimated based on 1 – r, where r is the pairwise Pearson's correlation between response patterns averaged across runs. Second, the **Classification Accuracy (CLA)** Dissimilarity was estimated based on the pairwise accuracy of a linear discriminant analysis (LDA) classifier using leave-one-run-out cross-validation. Third, the **Euclidean Distance (EUC)** Dissimilarity was estimated based on the (leave-one-run-out) cross-validated Euclidean distance: the pairwise distance between responses in the training data was multiplied by the transposed distances from the test data. Finally, the **Malahanobis Distance (MAL)** Dissimilarity was estimated based on the (leave-one-run-out) cross-validated Malahanobis distance: pairwise distances between training data responses were multiplied by the training data covariance, and transposed distances of the test data.

Noise Normalization

Following Walther et al., the data of individual subjects was normalized using the run-wise covariance between voxels as estimated by the GLM residuals from the first level analysis. To render the covariance matrices invertible, they were regularized using a shrinkage function.

Estimating Reliability and Effect Sizes

To estimate the within-subject reliability, for each subject data was split into odd and even runs, single RDMs (for each dissimilarity measure) were constructed for each split and correlated together (Pearson's r). For between-subject reliability, for each subject a single RDM (for each dissimilarity measure) was constructed, the RDM of one subject was correlated with the average of the remaining subjects, and then values were averaged across all leave-one-subject-out folds.

Neural RDMs of individual subjects were correlated with one of three behavioral or model RDMs: for Dataset 1, pairwise similarity judgments based on stimulus orientation and spatial frequency; For Dataset 2, multiple-arrangement similarity based on object shape similarity; and For Dataset 3, a binary model matrix differentiating social vs. non-social videos.

In order to assess the impact of noise normalization on the within- and between-subject reliability of RDMs, and correlation effects, a linear mixed effects model was fit to the data with normalization and measure as fixed effects, and subject as a random grouping effect.

Results

Within-Subject Reliability

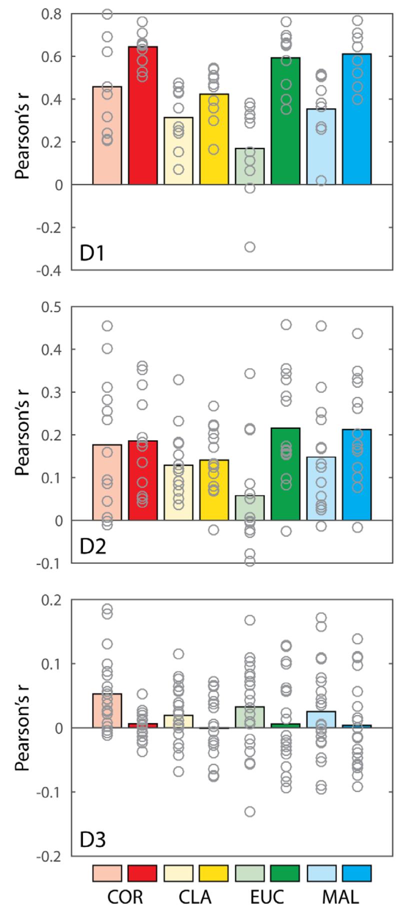

Figure 1: Mean within-subject reliability. Opaque colors indicate noise normalization, while faded colors indicate results without noise normalization

Normalization had a significant effect on the reliability for D1 ($F(1,72) = 19.36$, p = 3.67e-5). There was also a significant effect of measure ($F(3,72) = 11.63$, p = 2.66e-6), and a significant interaction between the two ($F(3,72) = 10.72$, p = 6.56e-6). When visualized (Figure 1), one can see that across measures noise normalization improve within-subject reliability, but the size of this improvement varies with measure, which also vary in baseline reliability.

In contrast to D1, there was no effect of noise normalization for D2 ($F(1,104) = .15$, p = .7), but there was a significant effect of measure (F(3,104) = 17.23, p = 3.48e-9), and interaction (F(3,104) = 16.36, p = 8.77e-9). For D2, it appears that normalization improves reliability for only some measures (Figure 2).

For D3, there was a consistent effect of noise normalization (F(1,160) = 11.39, p = 9.25e-4), but it was in the wrong direction (Figure 2). Although within-subject reliability was low for D3, it was made lower by normalizing. There were no significant effects of measure (F(3,160) = 1.42, p = .24), or interaction (F(3,160) = .8, p = .5).

Between-Subject Reliability

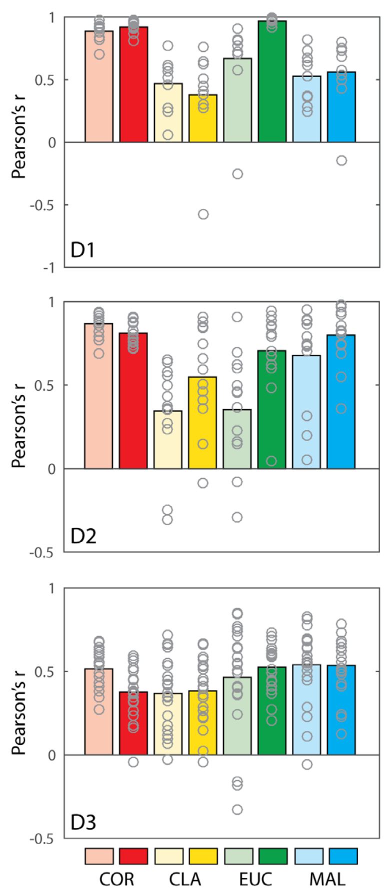

Figure 2: Mean between-subject reliability. Colors as in Figure 1.

When assessing between subject reliability, there was no effect of normalization (F(1,72) = .18, p = .67), or measure (F(3,72) = 2.6, p = .06), for D1. There was however a significant interaction between the two (F(,3,72) = 4.35, p = .007). Visual inspection suggests these results be influenced by some extreme outliers, whose data negatively correlated with that of the remaining subjects (Figure 2).

For D2, there was also no overall effect of normalization (F(1,104) = 1.02, p = .31), but there was an effect of measure (F(3,104) = 20.61, p = 1.48e-10), and an interaction between the two factors (F(3,104) = 9.28, p = 1.71e-5). Similar to D1, apparent differences in the mean between-subject reliability may be driven by extreme outliers (Figure 2).

For D3, there was an effect of normalization (F(1,160) = 6.73, p = .01), however this was again in the wrong direction. There were no effects of measure (F(1,160) = 2.52, p = .06), or significant interaction (F(3,160) = 2.58, p = .056). As with D1 and D2, apparent shifts in the group mean may be influenced by outliers.

Correlation Effects

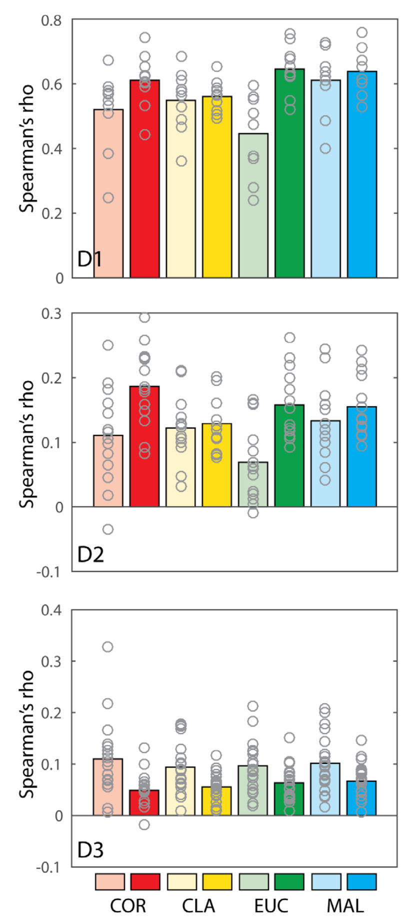

Figure 3: Mean RDM correlation. Colors as in Figure 1.

For D1, there were effects of normalization (F(1,72) = 23.29, p = 7.57e-6) and measure (F(3,72) = 9.67e-16) on the correlations of individual subject data with the behavioral RDM, as well as a significant interaction ($F(3, 72) = 34.72$, $p = 5.45e-14$). These results suggest that overall noise normalization increased the behavioral RDM correlations for D1 (Figure 3).

As with D1, we found effects of normalization ($F(1,104) = 49.13$, $p = 2.48e-10$), and measure ($F(3,104) = 21.89$, $p = 4.53e-11$), as well as a significant interaction ($F(3,104) = 21.8$, $p = 4.98e-11$). These results suggest that normalization also increased the behavioral RDM correlations for D2 (Figure 3).

Finally, for D3 we saw the same pattern once more, with effects of normalization ($F(1,160) = 72.13$, $p = 1.31e-14$), and measure ($F(3,160) = 5.24$, $p = .002$), and a significant interaction ($F(3,160) = 7.34$, $p = 1.22e-4$). However, noise normalization actually decreased the correlations with the binary model RDM (Figure 3).

Discussion

Noise normalization has been billed as a useful method for improving the reliability of neural RDMs (Walther et al., 2016). We revisited this issues to assess the extent to which noise normalization provides a net benefit to within-subject reliability of neural RDMs, while also assessing its impact on between-subject reliability, and correlations with behavioral or model RDMs.

Our findings were mixed. For only 1/3 datasets did we see a clear positive impact of noise normalization on within-subject reliability, while for 1/3 datasets reliability decreased after normalization. Across all three datasets, there was either no effect, or no positive effect, of noise normalization on between-subject reliability, which is typically used as the estimate of the noise ceiling for RSA. Finally, for 2/3 datasets noise normalization did seem to somewhat improve the correlations of neural RDMs with behavioral ones, while again for one dataset it made the correlations worse.

Taken as a whole, our findings provide equivocal support for making noise normalization core to the RSA pipeline. It is notable that for D3, noise normalization consistently made both within-subject reliability and model RDM correlations worse. While these findings could be related to the fact that there were far fewer runs than for D1 and D2 (6 compared to 12 and 16), they may also be related to the fact that the ROI was TPJ, which lacks the sort of topographic organization known to exist in the ROIs for the other datasets (object-LOTC, and especially V1). Therefore, one possibility is that methods like noise normalization may only provide a benefit when regions are known to have some clear topographic organization.

Acknowledgments

This work was supported by the European Research Council (ERC-2011-StG-284101), a Hercules grant ZW11_10, the Fund for Scientific Research-Flanders (G072517N), and the KU Leuven Research Council (C14/16/031), to H.O. S.B. was funded by FWO (Fonds Wetenschappelijk Onderzoek) through a postdoctoral fellowship (12S1317N) and a Research Grant (1505518N). This project has received funding from the FWO and European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665501, via a FWO [PEGASUS][2] Marie Skłodowska-Curie fellowship (12T9217N) to J.B.R.

References

Bracci, S., & Op de Beeck, H. (2016). Dissociations and associations between shape and category representations in the two visual pathways. *Journal of Neuroscience*, *36*(2), 432–444.

Lee Masson, H., Van De Plas, S., Daniels, N., & Op de Beeck, H. (2018). The multidimensional representational space of observed socio-affective touch experiences. *NeuroImage*, *175*, 297–314.

Ritchie, J. B., & Op de Beeck, H. (2019). A varying role for abstraction in models of category learning constructed from neural representations in early visual cortex. *Journal of Cognitive Neuroscience*, *31*(1), 155–173.

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*, 188–200.