Maximum Variance Differentiation (MVD) explains the transformation from IT to Perirhinal cortex

Marino Pagan[1], Eero P. Simoncelli[2,3], Nicole C. Rust[1]
[1]Dept. of Psychology, Univ. of Pennsylvania, [2]Center for Neural Science, New York Univ., [3]HHMI

Neural processing of signals for object recognition and target search has been shown to implement an "untangling" transformation, whereby initial population representations are converted into a format that is linearly separable. Despite the appeal of this description, understanding the precise nature of these computations has proven difficult. Here we propose that a transformation analogous to the well-known Energy model for V1 complex cells[1] can explain the untangling of target-match signals flowing from IT to Perirhinal cortex[2], collected as monkeys performed a delayed-match-to-sample task. For any N-way classification problem, "untangling" amounts to increasing the separation between the class means of the population response. For the IT to Perirhinal transformation, we hypothesized that untangling is achieved by transforming variance differences into mean differences through the use of a squaring operation. We optimized a linear-nonlinear-linear (LNL) response model to achieve this goal. The first linear stage transforms a population of IT inputs to maximize the variance differences between the classes in the output population. These linear responses are squared, and followed by a final orthogonal linear transformation. We find that a linear decoder operating on the responses of this Maximum Variance Differentiation (MVD) model attains target match vs. distractor performance close to that of an ideal observer operating on the IT population, and far better than a linear decoder operating directly on the input IT population, suggesting that most of the target match information embedded in IT population responses lies in the class variances. Furthermore, the MVD population matched Perirhinal linear decoder performance, suggesting that an MVD transformation within Perirhinal cortex may act on input arriving from IT. These results provide evidence that within the family of LNL models, a generalization of the Energy model is sufficient to explain the untangling of visual target match information.

1. Adelson EH and Bergen JR (1985) J. Opt. Soc. Am. A 2:284-299.
2. Pagan M, Urban LS, Wohl MP and Rust NC (2013) Nature Neurosci. 16:1132-39.

Methods: Populations of IT and Perirhinal neurons were recorded during a delayed-match-to-sample task in which monkeys had to indicate when a target image appeared in a sequence of distractors. Our previous report[2] described the untangling of target match versus distractor signals (a two-way classification) from IT to Perirhinal cortex. Here, we use an LNL generalization of the Energy model in which each model output neuron is created as a weighted sum of all the input neurons, followed by a squaring nonlinearity, and a final linear combination. We fit our model using a cross-validated procedure in which we used a subset of the trials from each condition to perform the fits, and we then tested population performance on separately measured trials. The model fitting procedure began by computing the optimal linear discriminant and eliminating it by projecting it out of the input population. From the remaining data, we estimated the linear weights for each output neuron by computing an eigendecomposition of the differences between target- and distractor-conditioned covariance matrices and we applied a squaring nonlinearity to these linear responses. Finally, because this procedure concentrated target match information in a small number of output neurons, we redistributed this information across the output population by applying a randomly generated orthogonal matrix. This solution is more direct, with a more intuitive and canonical model format that that of our previous work[2,] which used brute force search to optimize a more complex LN untangling model.

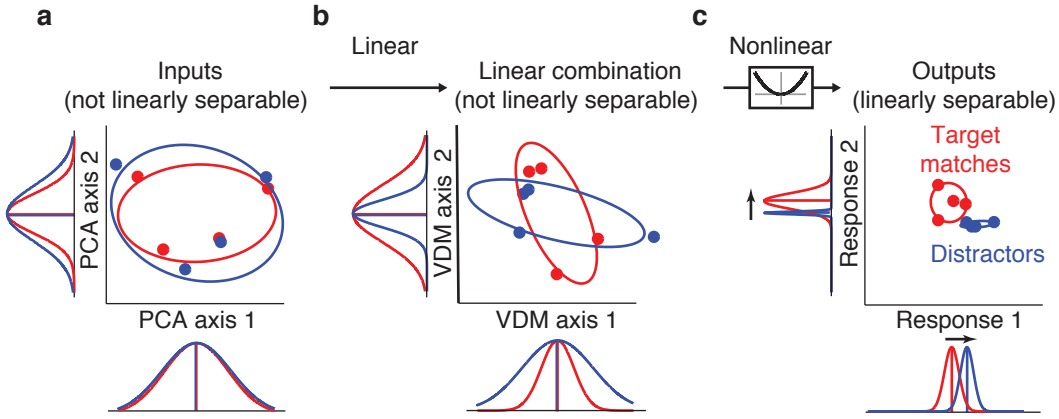

Figure 1. *MVD provides a normative account of how IT target match information might be untangled by perirhinal cortex.* In a delayed-match-to-sample task, monkeys viewed different images, and were required to indicate when a target match (red) image appeared within a sequence of distractors (blue). **a)** Projections of IT population responses onto the first two principle components, after removing the linear component of the task information. Note that the match and distractor conditions are not linearly separable in this space, and that the means and the variances of the marginal response distributions are approximately matched. **b)** The MVD transformation first rotates the input space to maximize differences between variances in the responses to matches and distractors. Shown are the responses of the first two components of this transformation. Note that the means of the distributions remain matched, and thus the representation is still not linearly separable. **c)** Output responses after applying the squaring nonlinearity, which essentially serves to translate variances into means, resulting in a more linearly separable (or untangled) population representation. The final (orthogonal) linear transformation is not shown.

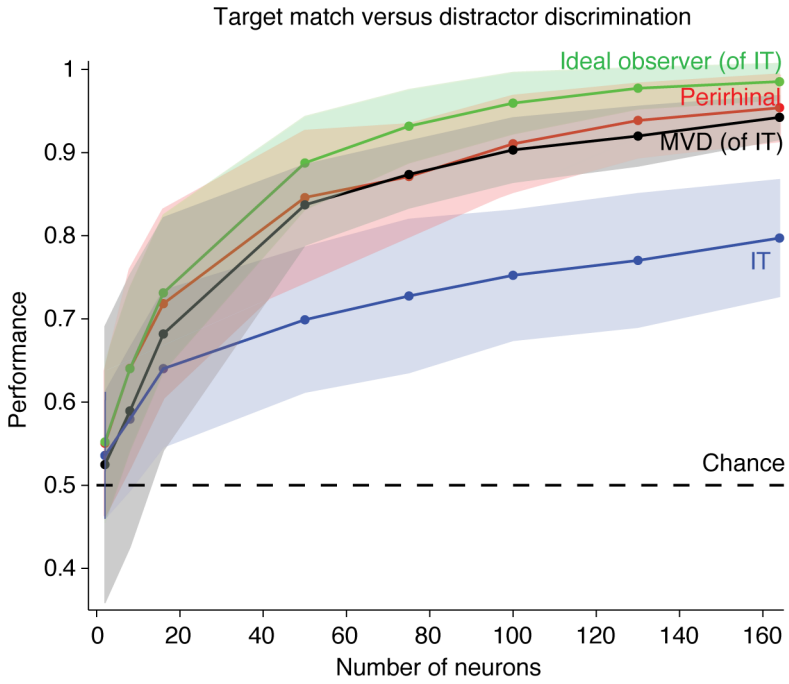

Figure 2. *MVD provides an accurate account of the untangling of IT target match information by Perirhinal cortex.* Cross-validated performance of a linear classifier for the target-distractor identification task, applied to the original IT population responses, the Perirhinal population responses, and the responses of the MVD model operating on the IT population. Also shown is performance of an ideal observer operating on the IT population.