The Neural Dynamics of Visual Processing in Monkey Extrastriate Cortex: A Comparison between Univariate and Multivariate Techniques

Maxime Cauchoix[1,2], Ali Bilgin Arslan[3], Denis Fize[1,2], and Thomas Serre[3]

[1] CNRS, CerCo, Faculté de médecine de Purpan, 31062, Toulouse, France
[2] Université de Toulouse, UPS, CerCo, 31062 Toulouse, France
[3] Department of Cognitive, Linguistic and Psychological Sciences, Brown Institute for Brain Science, Brown University, Providence RI 02912, USA

Abstract. Understanding the brain mechanisms underlying invariant visual recognition has remained a central tenet of cognitive neuroscience. Much of our current understanding of this process is based on knowledge gained from visual areas studied individually. Previous electrophysiology studies have emphasized the role of the ventral stream of the visual cortex in shape processing and, in particular, of higher level visual areas in encoding abstract category information. Surprisingly, relatively little is known about the precise dynamics of visual processing along the ventral stream of the visual cortex. Here we recorded intracranial field potentials (IFPs) from multiple intermediate areas of the ventral stream of the visual cortex in two behaving monkeys engaged in a rapid face categorization task. Using multivariate pattern analysis (MVPA) techniques, we quantified at millisecond precision the face category information conveyed by IFPs in areas of the ventral stream. We further investigate the relationship between the selectivity and latency of individual electrodes as estimated with classical univariate vs. multivariate techniques and conclude on the similarity and differences between the two approaches.

1 Introduction

Object recognition in primates is mediated by the ventral visual pathway of the visual cortex. This pathway runs from the primary visual cortex (V1) through extrastriate visual areas II (V2) and IV (V4), to the inferotemporal cortex (IT) (which can be further subdivided in a posterior (PIT) and anterior (AIT) part), and then to the prefrontal cortex (PFC), which is involved in linking perception to memory and action. A widely held view is that recognition is achieved through computations that gradually increase both the selectivity and the invariance properties of the underlying visual representation at each successive stages of the ventral stream of the visual cortex [16].

Much of our understanding of the mechanisms underlying visual recognition has, however, focused on these areas studied individually. Previous electrophysiology and imaging studies in both humans and monkeys have emphasized the role of the ventral stream of the visual cortex in shape processing [13,21] and,

in particular, of higher areas such as the IT cortex [12,18,17] and the PFC [4,5] for the coding of abstract category information. Specifically, these studies have shown that category information can be decoded in these areas within times that are fast enough to be consistent with behavioral responses [8,10,11]. Relatively little is known about the contribution of lower and intermediate visual areas to the categorization process and the underlying dynamics in these areas.

Here we use a rapid visual presentation paradigm [19] to investigate the neural basis of face processing in the ventral stream of the monkey visual cortex. We recorded intracranial field potentials (IFPs) from intermediate areas of the ventral stream of the visual cortex (areas V2, V4 and PIT) from two monkeys while they were actively engaged in a rapid face categorization task. Using multivariate pattern analysis (MVPA) techniques, we quantified at millisecond precision the face category information conveyed by IFPs.

2   Methods

Two male rhesus macaques (M1 and M2, 14 year old) were chronically implanted with subdural macro-electrodes and performed a face vs. non-face categorization task using a go/no-go paradigm (12 / 16 electrodes in total for M1/ M2). Stimuli were presented very briefly at the center of the screen for 32 ms. A total of 80 images were organized in 4 visual subcategories (10 images each) for both targets (macaque, chimpanzee, human and other faces) and distractors (fruits, flowers, mugs, and car tires). The images were equalized in luminance and RMS contrast to prevent low level biases in the recordings. All training and experimentation procedures conformed to the French and European standards for the use of experimental animals; protocols were approved by the regional ethical committee.

Using EEGLab, data was down-sampled to 250 Hz and epoched (-50 ms – 200 ms around the each stimuli onset). Three epoch rejection criteria were also employed: voltage, i.e., the power for one electrode or across all electrodes exceeded 3 standard deviations from the mean value computed over the entire recording session. The latter criterion targeted motor artifacts. We also excluded any trials that contained a non-varying signal (i.e., zero variance) over a 10-ms time windows (i.e., due to amplifier saturation). The data was z-scored and weights across electrodes are thus interpretable.

Linear SVMs were trained for the classification of targets (faces) vs. distractors (objects) (C parameter optimized using a cross-validation procedure) for each monkey separately. Here we considered SVM classifiers trained directly on the average potential over each electrode computed over 10 ms-long sliding time windows. The training and testing splits included multiple repetitions of the stimuli collected over several days.

An initial analysis using training and test sets obtained from different recording days suggested that the neural signal was largely stable over the course of days (Figure 1). Based on these results, we pooled all recordings sessions together. Random splits were generated so that the identity of the stimuli did not
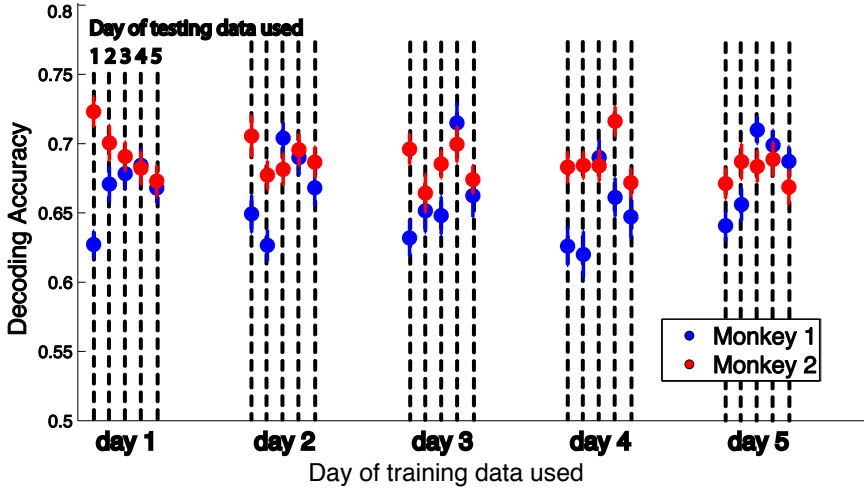
Fig. 1. Stability of the recorded signal for both monkeys: Decoding accuracy is reported for every possible training-testing set combination for different days

overlap between training and test (all repetitions of a given stimulus were thus used for either training or testing). This limits the possibility that the classifier simply "memorizes" patterns of neural activity associated with individual stimuli and instead pick on high-level category information.

Our analysis focuses on the initial 200 ms after stimulus onset to rule out any possible contamination from motor activity. We report the mean classification accuracy estimated over 50 cross-validations where the data was randomly split into two mutually exclusive halves for training and test (95% confidence intervals were computed based on the activity during the baseline period [-50ms 20ms]. These are indicated as vertical dashed lines around the chance level of decoding accuracy in Figures 2C and 4B). Two conditions are assumed to exhibit a statistically significant difference if the corresponding classifiers performance fall outside of these confidence intervals. This basic procedure was employed for all of decoding results, although the variables (i.e., electrodes included in the decoding) and the choice of decoding time points were manipulated for different types of analyses. Unless otherwise noted, classifiers are trained and tested on the same time points in the epoched and smoothed voltage data.

3   Results

The spatio-temporal nature of visual processing: The results of the neural decoding analysis are reported in Figure 2. Shown are the estimated latencies such that the decoding performance deviates significantly from baseline (60 ms for M1 and 88 ms for M2; see Methods). Overlaid on Figure 2A are the event related potentials (ERPs) for single electrodes. These ERPs, highlight both functional similarities within and differences between the three groups of electrodes (V2, V4

and PIT) thus confirming our initial anatomical localization. Figure 2B shows the corresponding weights of the linear SVM classifier averaged over each random split for every time point and expressed as a percentage of the maximum weight. These weights highlight the contribution of each electrode at specific time points and suggest that, at each time point, different areas and electrodes seem to encode the relevant category information. For M1, the contribution of electrodes over time seems to follow the known visual hierarchy (V2 → V4 → PIT).

Could the rapid decoding of category information from distributed patterns of neural data be predicted from univariate data analysis? Figure 2C shows a comparison between this multivariate analysis and similar analysis conducted on single electrodes (using the same classification procedure). The performance of a classifier trained on multiple electrodes distributed over areas V2, V4 and PIT exceeds the performance of the best electrode for every time point. This suggests that most of the time (for both monkeys, but more clearly for M2), electrodes in isolation do not provide nearly as much information as their combination. Overall this analysis reveals a very rapid signature for high-level category within 100 ms (in agreement with human electrophysiology recordings [11,20]), which is supported by a distributed representation over areas and electrodes and that evolves over time as the signal progresses through the visual hierarchy.

Univariate vs. multivariate pattern analysis: Figure 3A provides a comparison between classical univariate significance tests for each electrode, the time points with the corresponding weights acquired from the MVPA analysis, and the corresponding univariate decoding performance for this electrode using the same SVM classifier ($p < 0.01$ for all tests; correction for multiple comparisons is approximated by seeking consistent significance for at least ten consecutive time
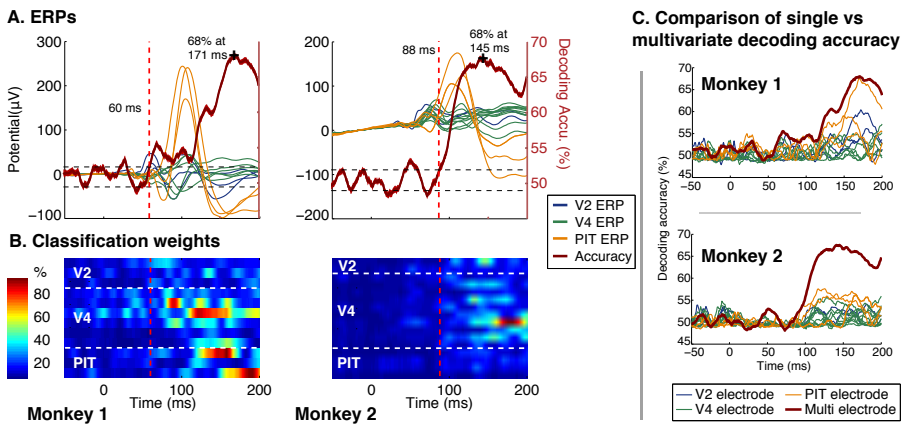

Fig. 2. A) Overall classification accuracy as a function of time; B) corresponding classification weights for each electrode and time point and C) decoding accuracy for single vs. multiple electrodes
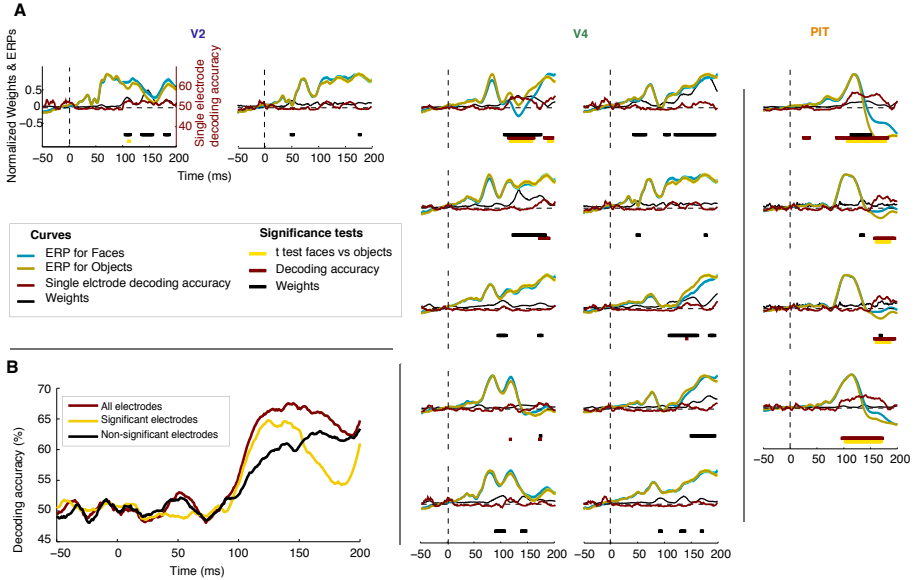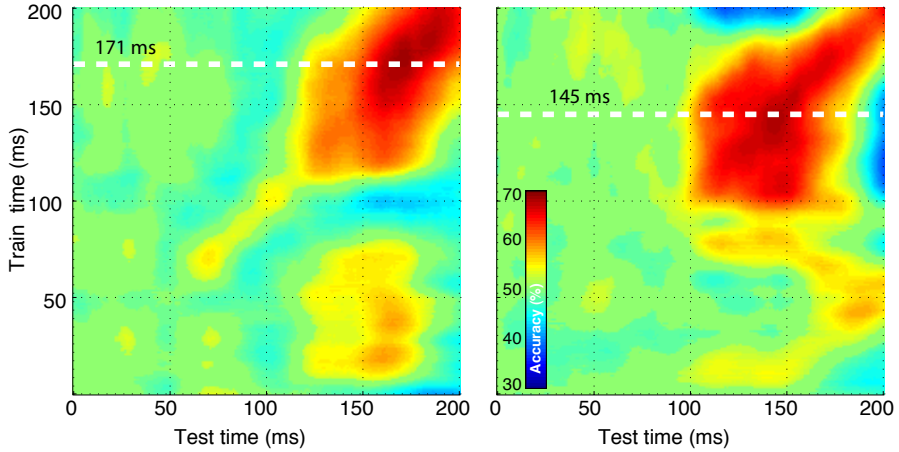
Fig. 3. Single versus multiple electrode analysis of informativeness measures for M2. A) ERP comparison between targets (cyan) and distractors (yellow) with decoding performance for the corresponding electrode (red) and weights resulting from MVPA (black). T-tests for targets vs. distractors are shown with bars. B) MVPA decoding performance using all vs. subsets of electrodes with significant (yellow lines) & non-significant differential activity.

points. This is a common practice employed by other researchers [19] and it also seems to be compatible with Bonferroni and FDR corrections. We chose this method for a more intuitive explanation). Interestingly, both t-test and univariate decoding exhibit a very similar trend. However, the electrode contributions assessed via MVPA seems to indicate qualitative diffxerences. For instance, for both monkeys, some electrodes (mainly in V4) seem to be associated with high classification weights although these electrodes do not seem to exhibit any significant differential activity as measured with classical univariate analysis. Overall this suggests that classical univariate analyses may underestimate the contributions of single electrodes. Last, Figure 3B shows that removing electrodes that exhibit significant differential activity tend to slow down the decoding of category information.

Characterizing the nature of the neural code. We also conducted a *discriminant cross-training* (DCT) analysis [1]: A classifier is trained using patterns of neural activity at one time point t_o and tested on all time points in the epoch, producing a matrix of decoding accuracy for each combination of training/testing time points (figure 4A). Note that the diagonals of these two matrices correspond

A. Discriminant cross-training

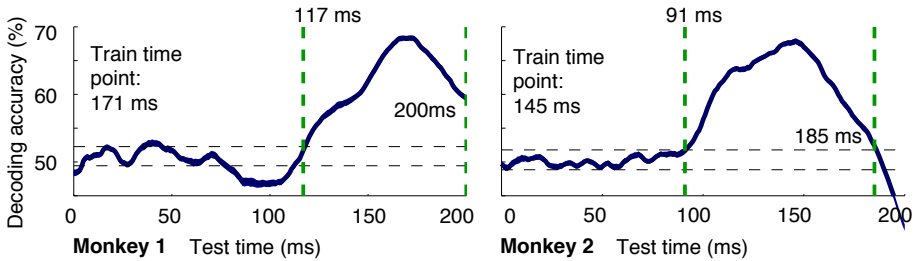**B. Duration of the template of maximum decoding**

Fig. 4. A) Discriminant cross-training (DCT) analysis: Color indicates decoding accuracy for each combination of training and test times (vertical and horizontal axes respectively). B) Decoding accuracy for a classifier trained on the time point where the overall decoding was maximal (dashed line, see figure 2), and tested on the remaining time points.

to the decoding accuracy curves shown in figure 2C. This analysis allows us to assess how suitable is the neural template picked up by the classifier over time. As seen on figure 4A, the neural template that emerges around 100 ms post-stimulus (X axis), and remains stable for about 100 ms, suggesting a sustained coding of facial information in this set of areas possibly supported by feedback. Figure 4B provides a cross-section taken from the upper graphs (on dashed white lines). The rationale behind taking these points in time to show the duration of the "neural template" was the assumption that category information would be maximal at these time points in the trial and therefore they would be the best approximations of temporally adjacent task relevant brain states.

4   Discussion

Psychophysical studies using the rapid categorization paradigm [19], have demonstrated the remarkable speed at which the visual system can operate (with reac-

tion times as fast as 200 ms for monkeys and 250 ms for human participants [3]). The present study provides a quantitative description of the signal propagated through the ventral stream of the visual cortex during the task. The present analysis suggests latencies for the coding of face category that are shorter than previously estimated from single electrode IT recordings [9]. The very short latencies reported in the present study are, however, compatible with those estimated from human electrophysiology [11,20] and psychophysics experiments [2,7]. These results may help shed some light on a long-standing debate on the role played by various EEG/MEG components in face perception. However it is important to emphasize that care should be taken in interpreting these latencies as some of these differences may be influenced by the task difficulty and the stimulus set used. The present decoding analysis may highlight a monkey homologue of the P100 assumed to reveal one of the first stages of face processing [6,15] starting in the Occipital Face Area (OFA).

The present study also demonstrates the stability of the neural signal over days of recording and the ability of a linear classifier to genralize over recording session. Furthermore our analysis highlights some of the key benefits of MVPA as compared to a classical ERP/univariate analysis by suggesting that intermediate visual areas may contribute relatively early coding signals for the processing of faces. This idea seems compatible with the high level of performance obtained with low-level visual features for the processing of faces [14]. Overall this suggests that intermediate visual areas such as V4 may already play a key role in visual recognition and that the coding of category information may be implemented via a distributed process in both space and time.

References

1. Carlson, T.A., Hogendoorn, H., Kanai, R., Mesik, J., Turret, J.: High temporal resolution decoding of object position and category. Journal of Vision 11(10), 9, 1–17 (2011)
2. Crouzet, S.M., Kircher, H., Thorpe, S.J.: Fast saccades toward faces: Face detection in just 100 ms. Journal of Vision 10(4), 1–17 (2010)
3. Fize, D., Cauchoix, M., Fabre-Thorpe, M.: Humans and monkeys share visual representations. Proceedings of the National Academy of Sciences of the United States of America 108(18), 7635–7640 (2011)
4. Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K.: Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291(5502), 312–316 (2001)

5. Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K.: A comparison of primate prefrontal and inferior temporal cortices during visual categorization. The Journal of Neuroscience 23(12), 5235–5246 (2003)
6. Freiwald, W.A., Tsao, D.Y.: Functional compartmentalization and viewpoint generalization within the macaque face-processing system. Science 330(6005), 845–851 (2010)
7. Girard, P., Jouffrais, C., Kirchner, C.H.: Ultra-rapid categorisation in non-human primates. Animal Cognition 11(3), 485–493 (2008)
8. Hung, C.P., Kreiman, G., Poggio, T., DiCarlo, J.J.: Fast readout of object identity from macaque inferior temporal cortex. Science 310(5749), 863–866 (2005)
9. Kiani, R., Esteky, H., Tanaka, K.: Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. Journal of Neurophysiology 94(2), 1587–1596 (2005)
10. Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Bandettini, P.A.: Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60(6), 1126–1141 (2008)
11. Liu, H., Madsen, J.R., Agam, Y., Kreiman, G.: Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. Neuron 62(2), 281–290 (2009)
12. Logothetis, N.K., Sheinberg, D.L.: Visual object recognition. Annual Review of Neuroscience 19(1), 577–621 (1996)
13. Pasupathy, A., Connor, C.E.: Population coding of shape in area V4. Nature Neuroscience 5(12), 1332–1338 (2002)
14. Pinto, N., DiCarlo, J.: How far can you get with a modern face recognition test set using only simple features? In: Vision and Pattern Recognition, pp. 2591–2598 (2009)
15. Pitcher, D., Walsh, V., Duchaine, B.: The role of the occipital face area in the cortical face perception network. Experimental Brain Research 209(4), 481–493 (2011)
16. Riesenhuber, M., Poggio, T.: Hierarchical models of object recognition in cortex. Nature Neuroscience 2(11), 1019–1025 (1999)
17. Sigala, N., Logothetis, N.K.: Visual categorization shapes feature selectivity in the primate temporal cortex. Nature 415(6869), 318–320 (2002)
18. Tanaka, K.: Inferotemporal cortex and object recognition. Annual Review of Neuroscience 19, 109–139 (1996)
19. Thorpe, S., Fize, D., Marlot, C.: Speed of processing in the human visual system. Letters to Nature 381, 520–522 (1996)
20. Tsuchiya, N., Kawasaki, H., Oya, H., Howard, M.A., Adolphs, R.: Decoding face information in time, frequency and space from direct intracranial recordings of the human brain. PloS One 3(12), e3892 (2008)
21. Ungerleider, L.G., Bell, A.H.: Uncovering the visual "alphabet": advances in our understanding of object perception. Vision Research 51(7), 782–799 (2011)