Supplementary Information

Neural face representations are mainly transient. To investigate whether the neural representations of face images were stable over time or continuously changing, we performed temporal generalization analysis [1-3]. In particular, we trained classifiers on data at a specific time point *t* and tested it on all other time points. If neural representations are stable, the classifier should generalize well across time points. We found that neural representations were sustained for a short amount of time (up to around 200 ms), and quickly decreased below chance for larger time spans (Fig. S1). Surprisingly, we found a lack of generalization very early between around 100 and 200 ms (shown as non-significant off-diagonal areas between 100 and 200 ms), with a reactivation of the neural representations at later stages. This could reflect some early feedback or recurrent processing. Overall, we found that neural representations are mainly transient and do not persist over time, suggesting continuous transformations of stimulus information.

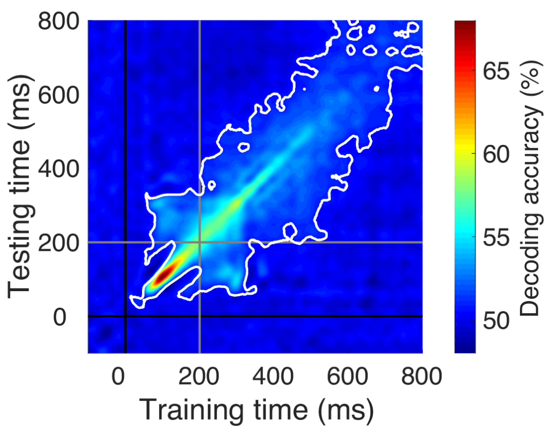

Figure S1 | Temporal generalization of face image decoding. The correlation classifier trained with principal components extracted from all MEG sensors at time point *t* (training time) and tested on all other time points (testing time). The temporal generalization decoding matrix was averaged over all image pairs and all subjects, thus corresponding to the temporal generalization of the image decoding time course shown in Fig. 2a. Black line marks image onset, and the gray line marks the image offset. The white contour indicates significant decoding values using one-sided cluster-based sign permutation test (cluster-defining threshold $p < 0.05$, and corrected significance level $p < 0.05$).

Temporal generalization of MEG multivariate pattern analysis. To get a measure of how well neural representations generalized across time, we extended the multivariate pattern analysis by a temporal generalization approach [1-3]. Based on this approach, a classifier is trained on the data of a specific time point *t* (training time) and tested on all other time points *t'* (testing time). Intuitively, if representations are stable over time, a classifier should be able to discriminate two conditions not only at the trained time, but also at later time points. We used the same classification scheme as used for classification on identical time points. However, to save computation time, we used a cross-validated pairwise correlation classifier instead of a linear support vector machine (SVM). This classifier reached slightly lower decoding accuracy values when trained on identical time points (i.e. the diagonal values in Fig. S1) than the SVM (c.f. Fig 2a) but a highly similar pattern over time. Similar to the classification analysis at identical time points, we performed temporal generalization analysis separately for each subject and averaged the decoding accuracies across stimulus pairs. This yielded a 901 x 901 matrix (-100 to 800 ms with respect to stimulus onset) for each subject.

References

1. King, J.-R. & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.* **18,** 203–210 (2014).
2. Isik, L., Meyers, E. M., Leibo, J. Z. & Poggio, T. The dynamics of invariant object recognition in the human visual system. *J. Neurophysiol.* **111,** 91–102 (2014).
3. Cichy, R. M., Pantazis, D. & Oliva, A. Resolving human object recognition in space and time. *Nat. Neurosci.* **17,** 1–10 (2014).