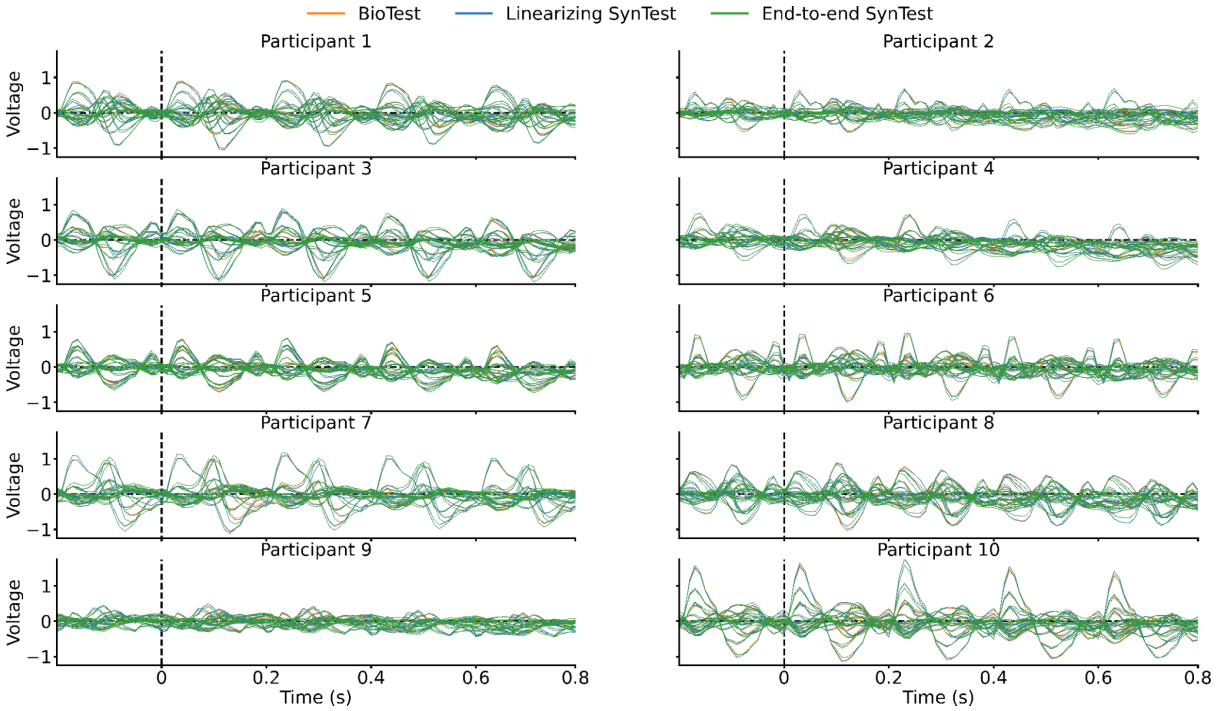
A large and rich EEG dataset for modeling human visual object recognition

Alessandro T. Gifford, Kshitij Dwivedi, Gemma Roig, Radoslaw M. Cichy

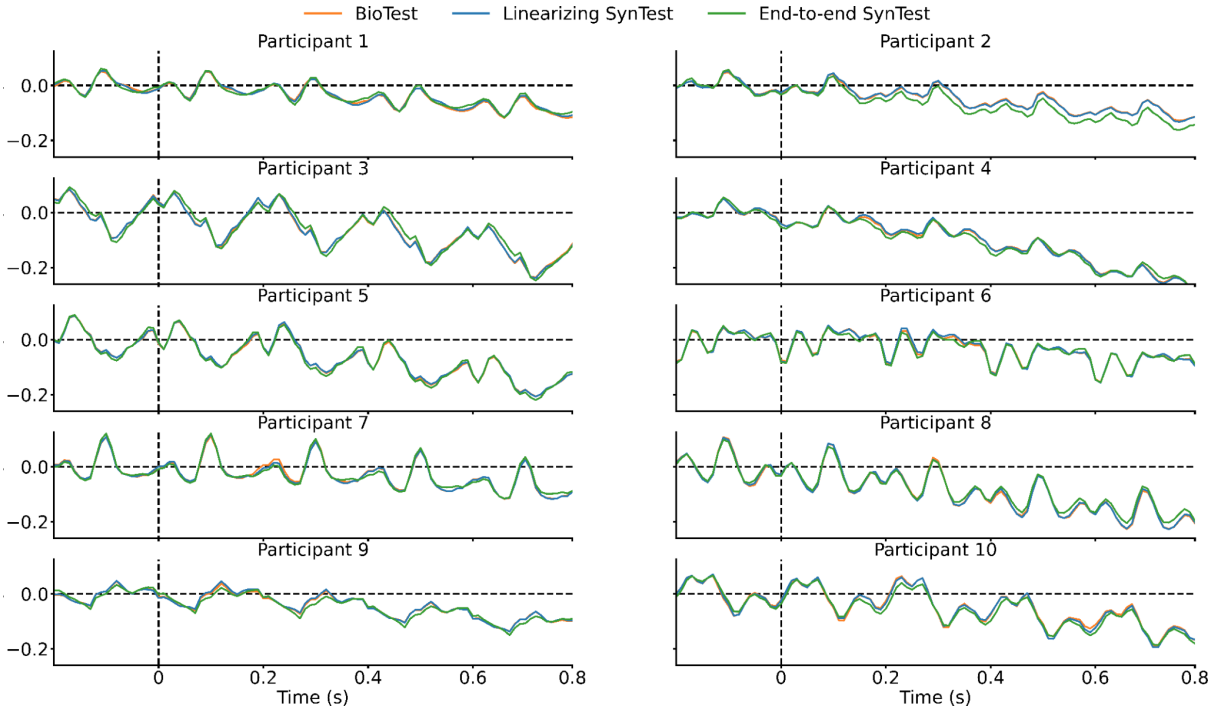
SUPPLEMENTARY MATERIAL

Supplementary Figure 1: ERPs of single EEG channels, single participants



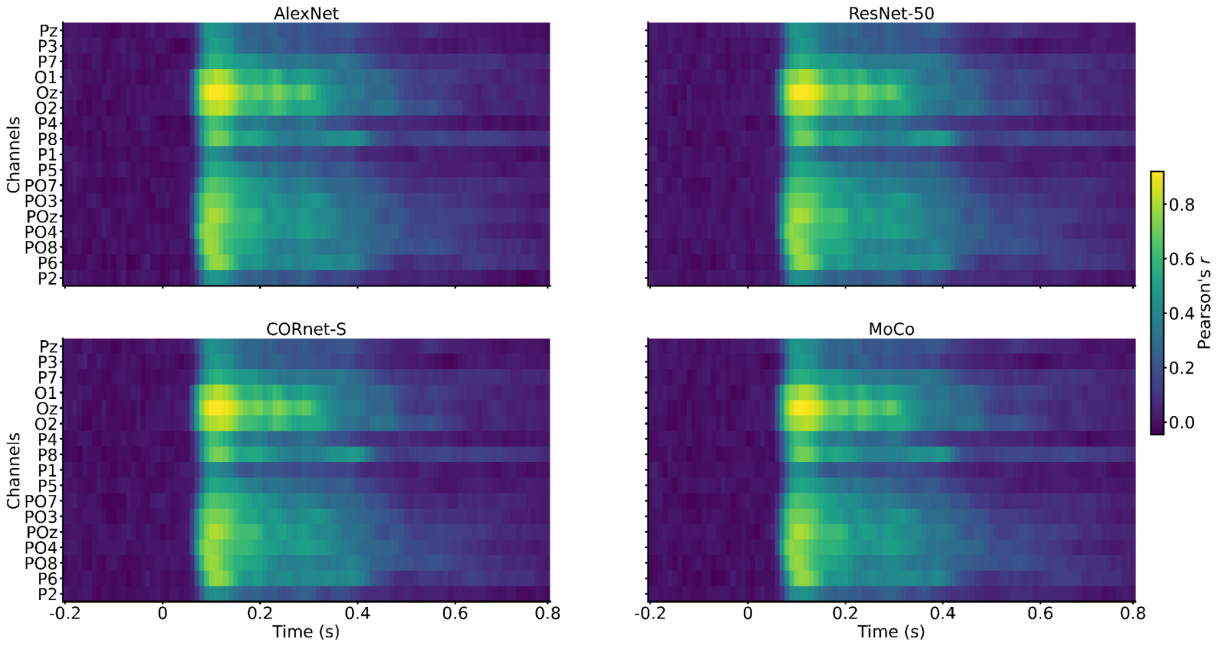
Supplementary Figure 1. Single EEG channels ERPs of the BioTest data, AlexNet linearizing SynTest data, and AlexNet end-to-end SynTest data (of the DNNs trained to predict all EEG time points at once), obtained by averaging the EEG signal across image conditions and repetitions. The ERPs of the biological and synthetic data are largely overlapping.

Supplementary Figure 2: Channels-average EEG ERPs, single participants



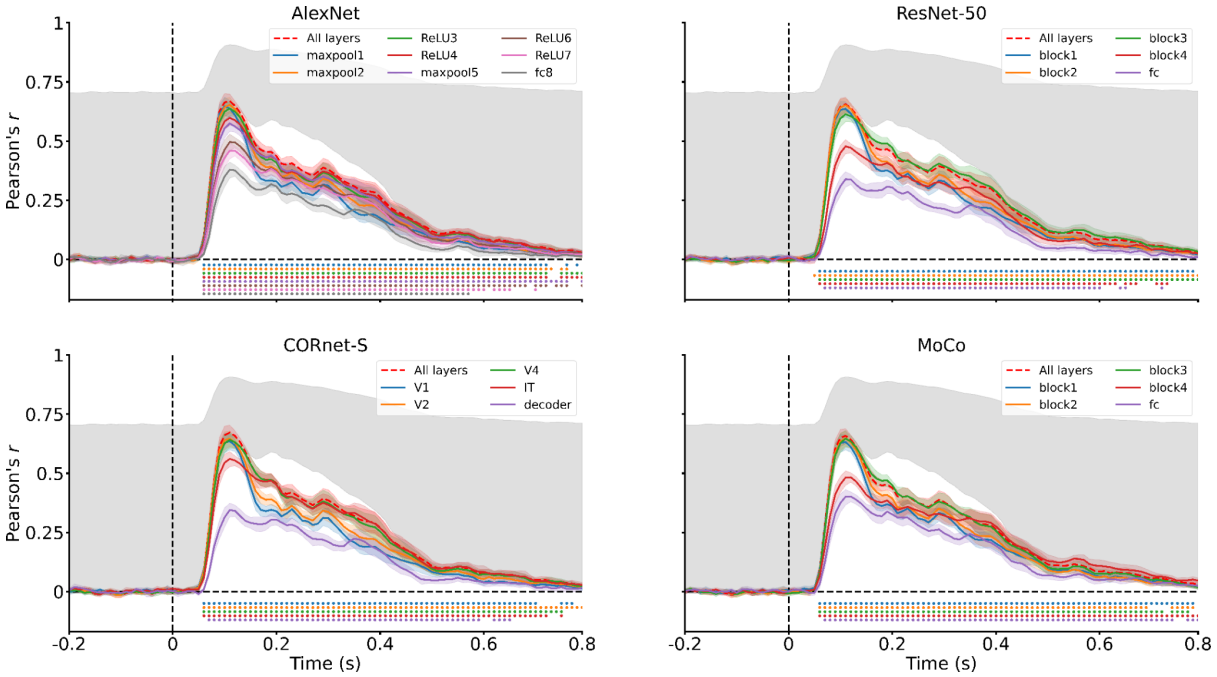
Supplementary Figure 2. Average EEG channels ERPs of the BioTest data, AlexNet linearizing SynTest data, and AlexNet end-to-end SynTest data (of the DNNs trained to predict all EEG time points at once), obtained by averaging the EEG signal across image conditions and repetitions. The ERPs of the biological and synthetic data are largely overlapping.

Supplementary Figure 3: Linearizing encoding correlation results, individual EEG channels



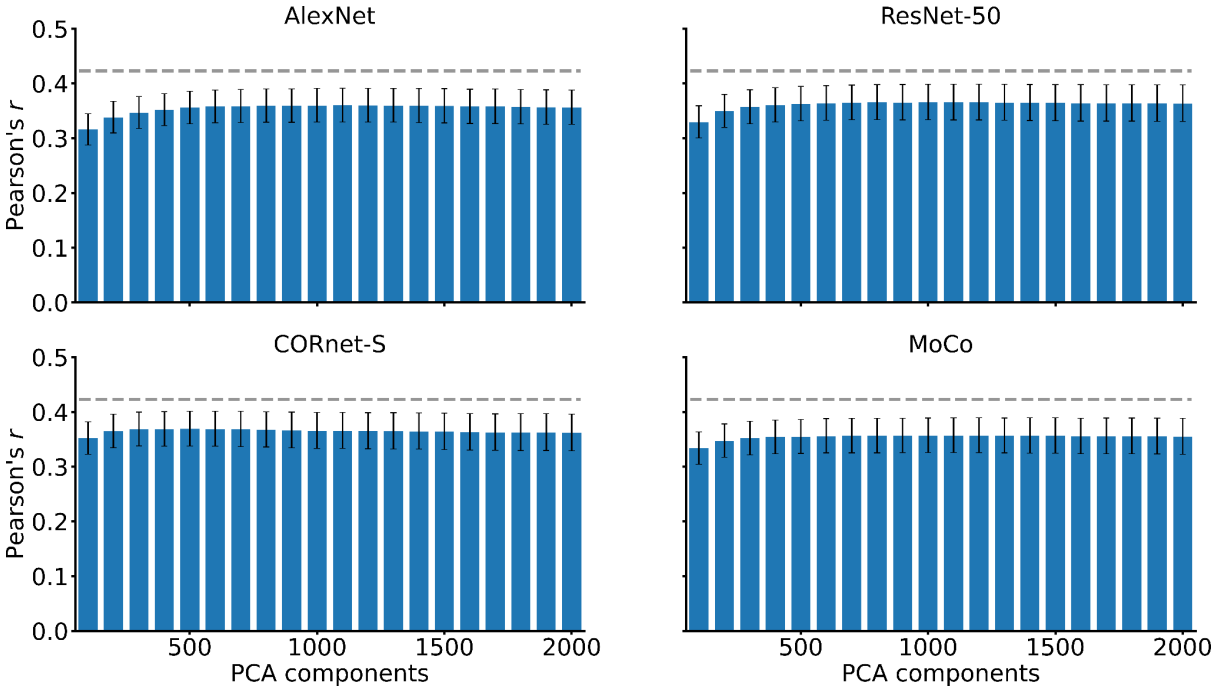
Supplementary Figure 3. Single EEG channels correlation results of the linearizing encoding models, averaged across participants. Occipital channels are best modeled, followed by parieto-occipital channels and finally parietal channels.

Supplementary Figure 4: Linearizing encoding with individual DNN layers



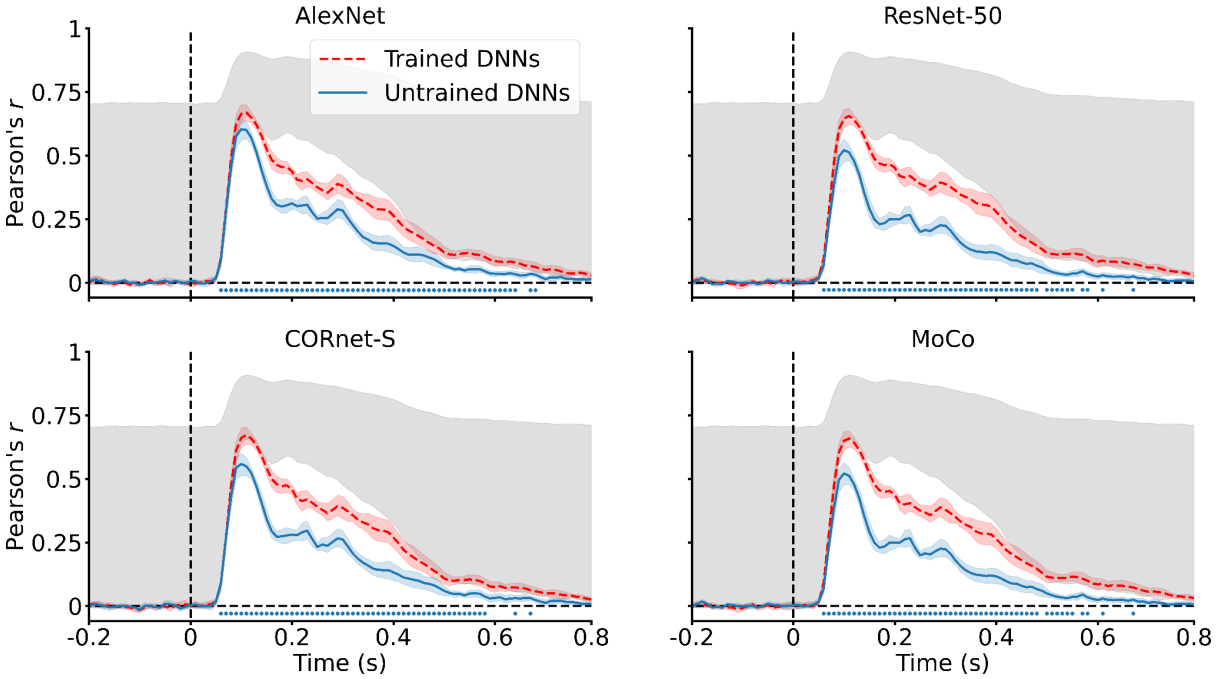
Supplementary Figure 4. We trained linearizing encoding models using the PCA downsampled feature maps of individual DNN layers, and evaluated them through the correlation analysis. Here we present the correlation results averaged across participants. Earlier parts of the EEG response (< 200ms) are better modeled by initial DNN layers, whereas later parts of the EEG response (> 200ms) are better modeled by intermediate/higher DNN layers. Error margins, asterisks, gray area and black dashed lines as in **Figure 4**.

Supplementary Figure 5: Linearizing encoding using DNN feature maps with different amounts of PCA components



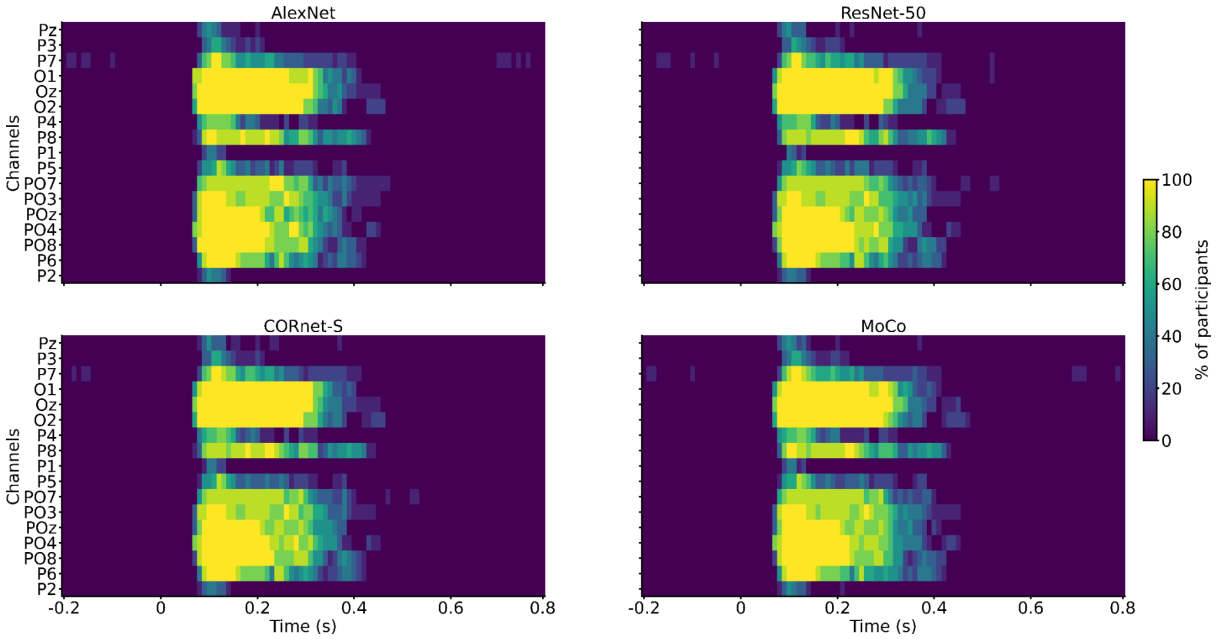
Supplementary Figure 5. We trained linearizing encoding models using DNN feature maps with different amounts of PCA components, and evaluated them through the correlation analysis. Here we present the correlation results averaged across participants. The prediction accuracies slightly increase in the range [100 500] PCA components, and are nearly identical in the range [500 2000] PCA components. Error margins, gray dashed lines and black dashed lines as in **Figure 7**.

Supplementary Figure 6: Linearizing encoding with feature maps of untrained DNNs



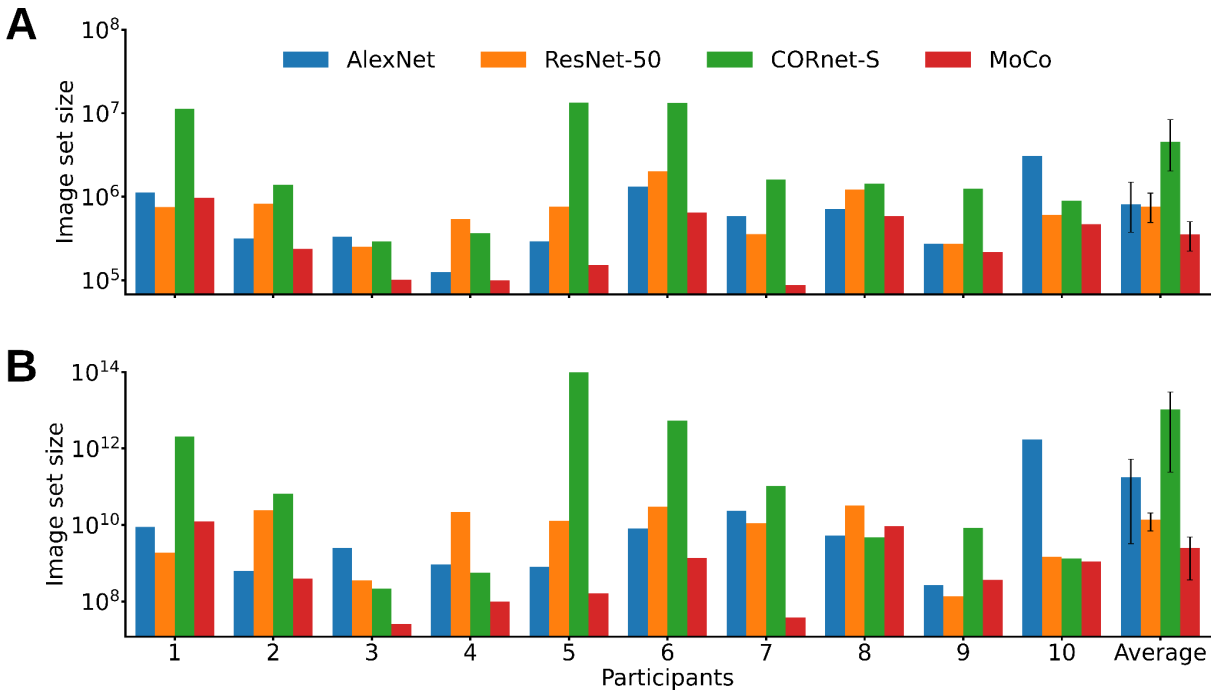
Supplementary Figure 6. We trained linearizing encoding models using feature maps of untrained DNNs, and evaluated them through the correlation analysis. Here we present the correlation results averaged across participants. Untrained networks explain a significant portion of variance. Error margins, asterisks, gray area and black dashed lines as in **Figure 4**.

Supplementary Figure 7: Zero-shot identification best EEG features



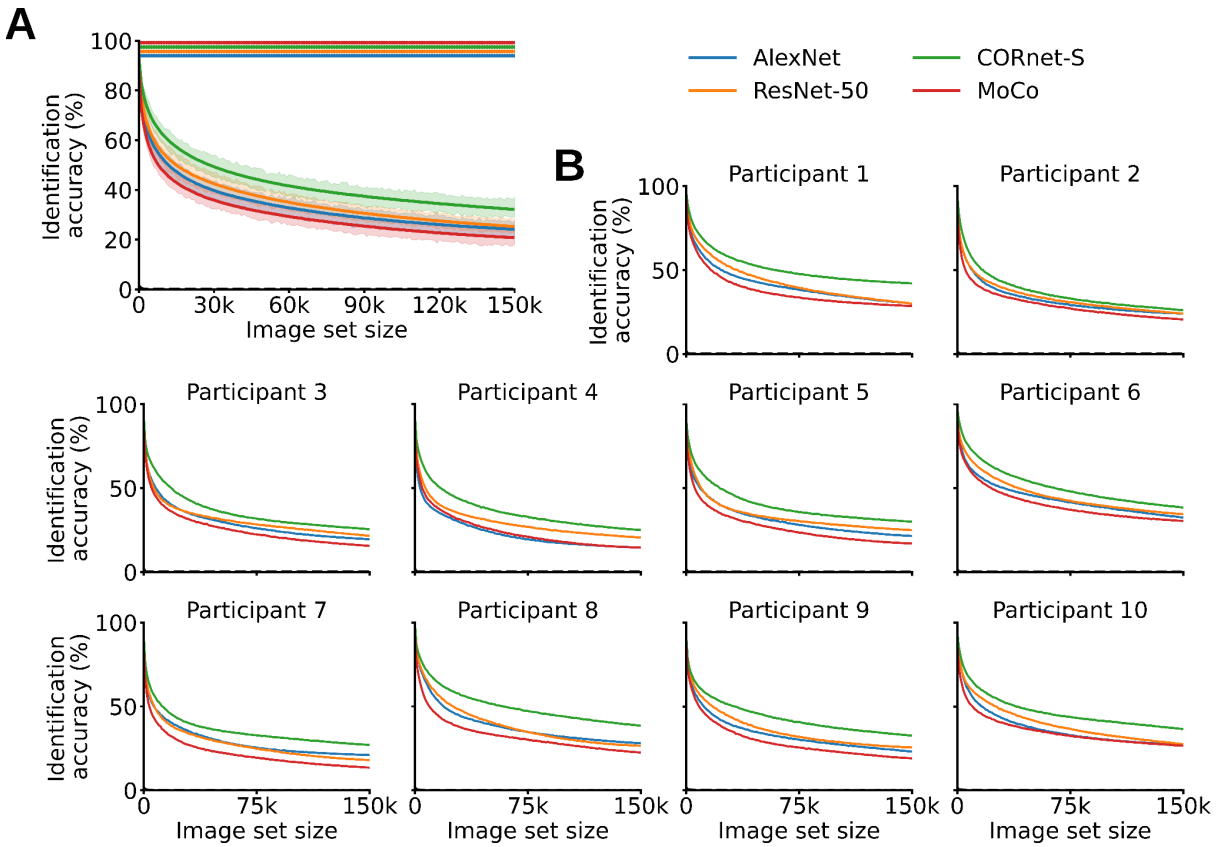
Supplementary Figure 7. 300 best EEG features used for the zero-shot identification analysis, summed over subjects. The best features are mostly occipital and parieto-occipital channels between 70ms and 400ms after stimulus onset.

Supplementary Figure 8: Zero-shot identification top-1 extrapolations



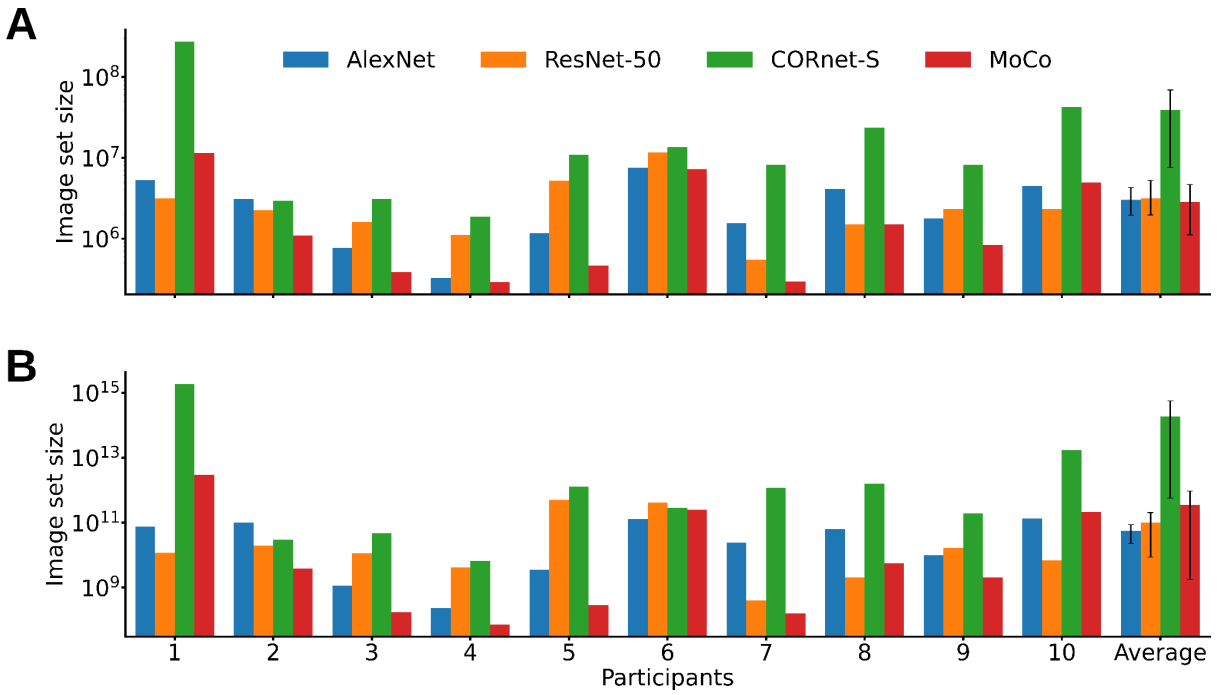
Supplementary Figure 8. Extrapolation of the zero-shot identification accuracy as a function of candidate image set sizes. (**A**) Candidate image set sizes required for the identification accuracy to drop to 10%. (**B**) Candidate image set sizes required for the identification accuracy to drop to 0.5%.

Supplementary Figure 9: Zero-shot identification top-3 results



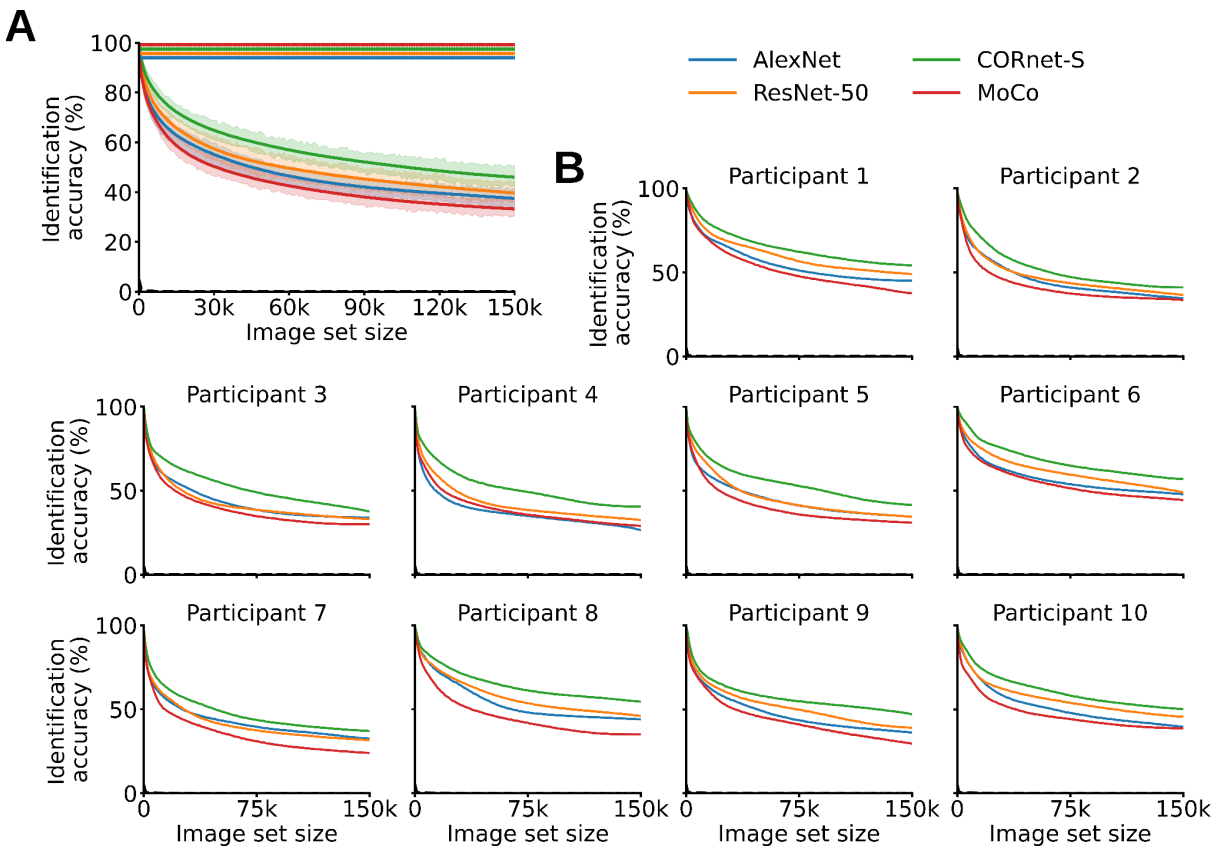
Supplementary Figure 9. Zero-shot identification, three most correlated candidate image conditions. Zero-shot identification of the BioTest data using the SynTest data and the synthesized EEG visual responses to the 150,000 ILSVRC-2012 validation and test image conditions (SynImagenet), with the correct image condition falling within the three most correlated image conditions. (**A**) Zero-shot identification results averaged across participants. With a SynImagenet set size of 0 the synthesized data of AlexNet, ResNet-50, CORnet-S, MoCo significantly identify the BioTest data with accuracies of, respectively, 90.3%, 90.95%, 93.7%, 88.25%. (P < 0.05, one-sample one-sided t-test, Bonferroni-corrected). With a SynImagenet set size of 150,000 the synthesized data of AlexNet, ResNet-50, CORnet-S, MoCo significantly identify the BioTest data with accuracies of, respectively, 24.15%, 25.30%, 32.15%, 20.80%. (**B**) Individual participants' results. Error margins and black dashed lines as in **Figure 4**. Asterisks as in **Figure 6**.

Supplementary Figure 10: Zero-shot identification top-3 extrapolations



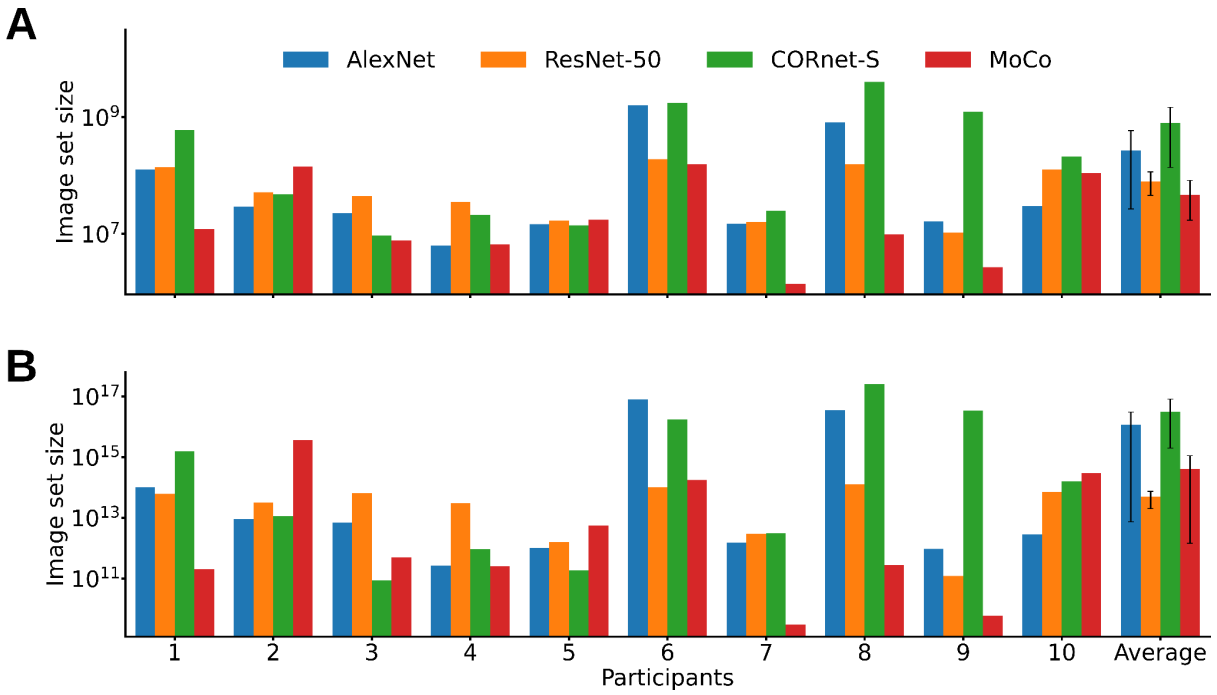
Supplementary Figure 10. Zero-shot identification, three most correlated candidate image conditions. Extrapolation of the zero-shot identification accuracy as a function of candidate image set sizes. (**A**) Candidate image set sizes required for the identification accuracy to drop to 10%. (**B**) Candidate image set sizes required for the identification accuracy to drop to 0.5%.

Supplementary Figure 11: Zero-shot identification top-10 results



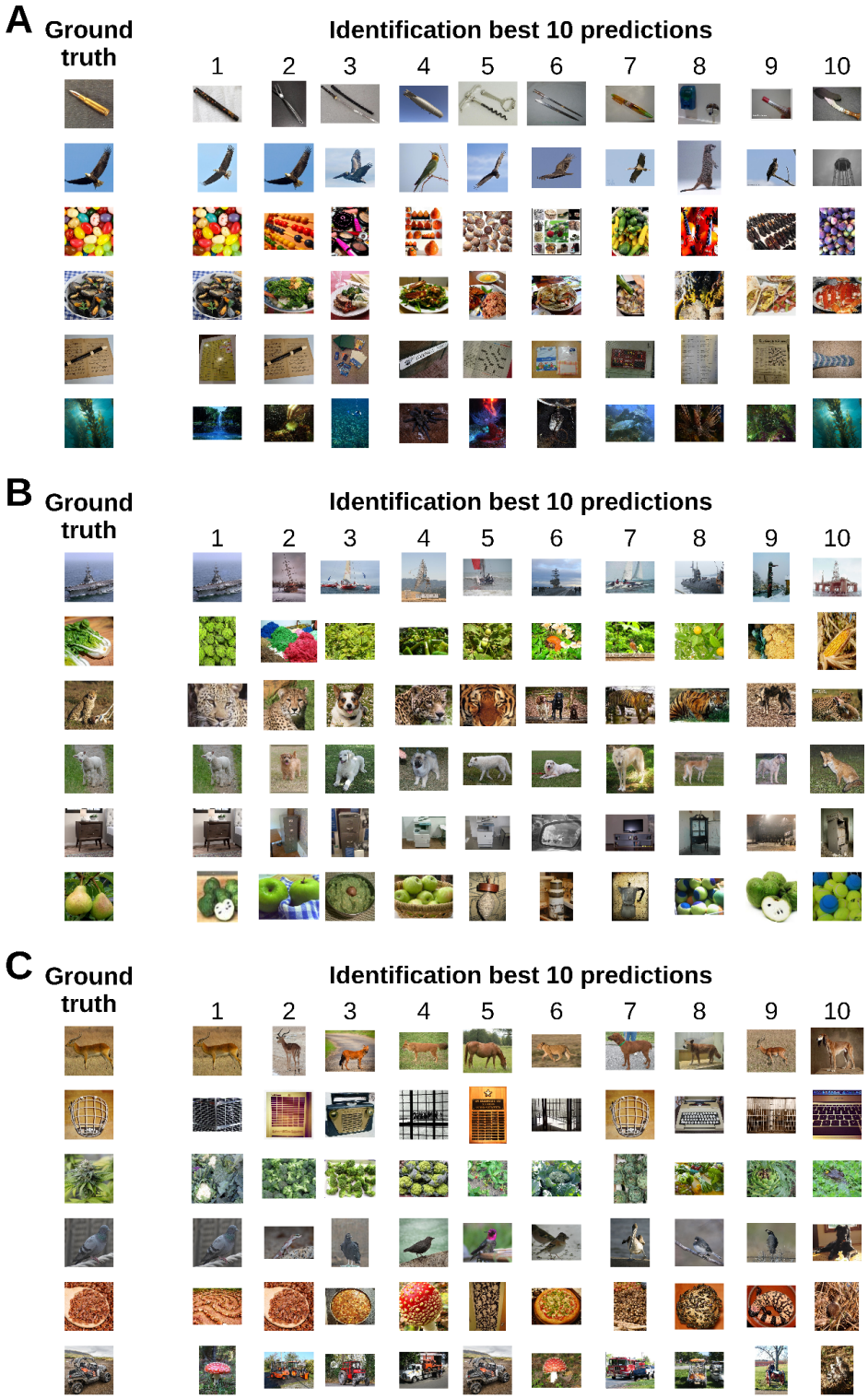
Supplementary Figure 11. Zero-shot identification, ten most correlated candidate image conditions. Zero-shot identification of the BioTest data using the SynTest data and the synthesized EEG visual responses to the 150,000 ILSVRC-2012 validation and test image conditions (SynImagenet), with the correct image condition falling within the ten most correlated image conditions. (**A**) Zero-shot identification results averaged across participants. With a SynImagenet set size of 0 the synthesized data of AlexNet, ResNet-50, CORnet-S, MoCo significantly identify the BioTest data with accuracies of, respectively, 97.55%, 97.80%, 99.20%, 97.05%. ($P < 0.05$, one-sample one-sided t-test, Bonferroni-corrected). With a SynImagenet set size of 150,000 the synthesized data of AlexNet, ResNet-50, CORnet-S, MoCo significantly identify the BioTest data with accuracies of, respectively, 37.45%, 39.65%, 46%, 33.25%. (**B**) Individual participants' results. Error margins and black dashed lines as in **Figure 4**. Asterisks as in **Figure 6**.

Supplementary Figure 12: Zero-shot identification top-10 extrapolations



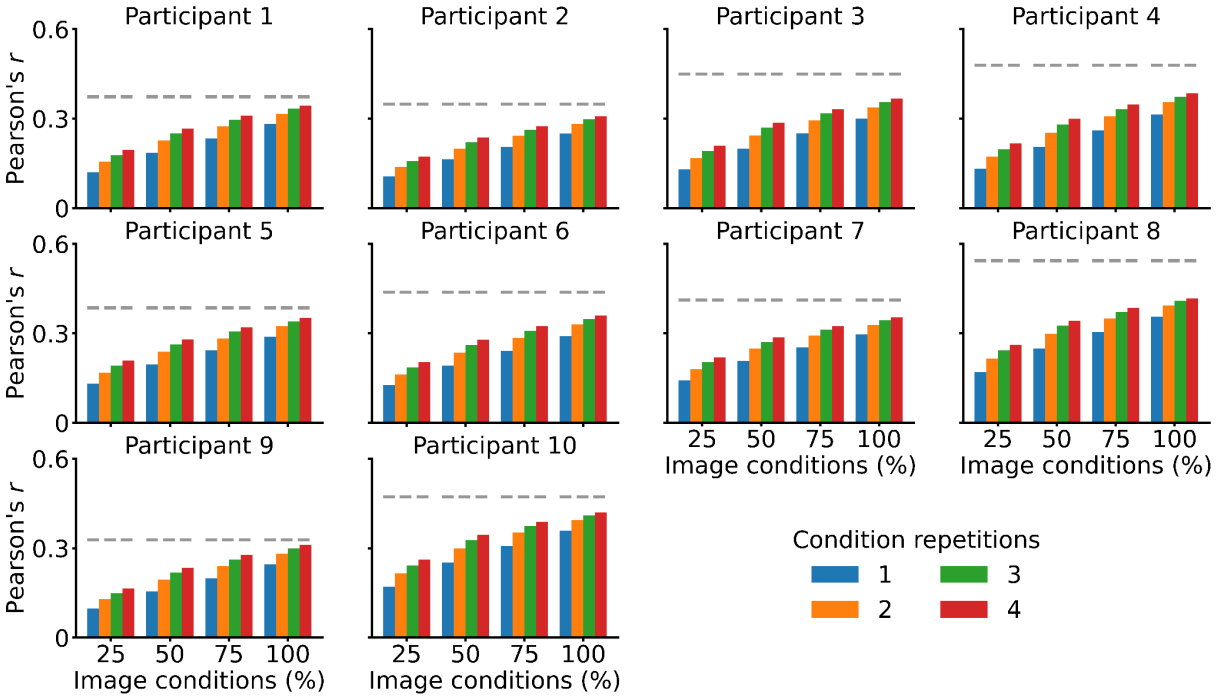
Supplementary Figure 12. Zero-shot identification, ten most correlated candidate image conditions. Extrapolation of the zero-shot identification accuracy as a function of candidate image set sizes. (**A**) Candidate image set sizes required for the identification accuracy to drop to 10%. (**B**) Candidate image set sizes required for the identification accuracy to drop to 0.5%.

Supplementary Figure 13: Zero-shot identification predictions visualizations



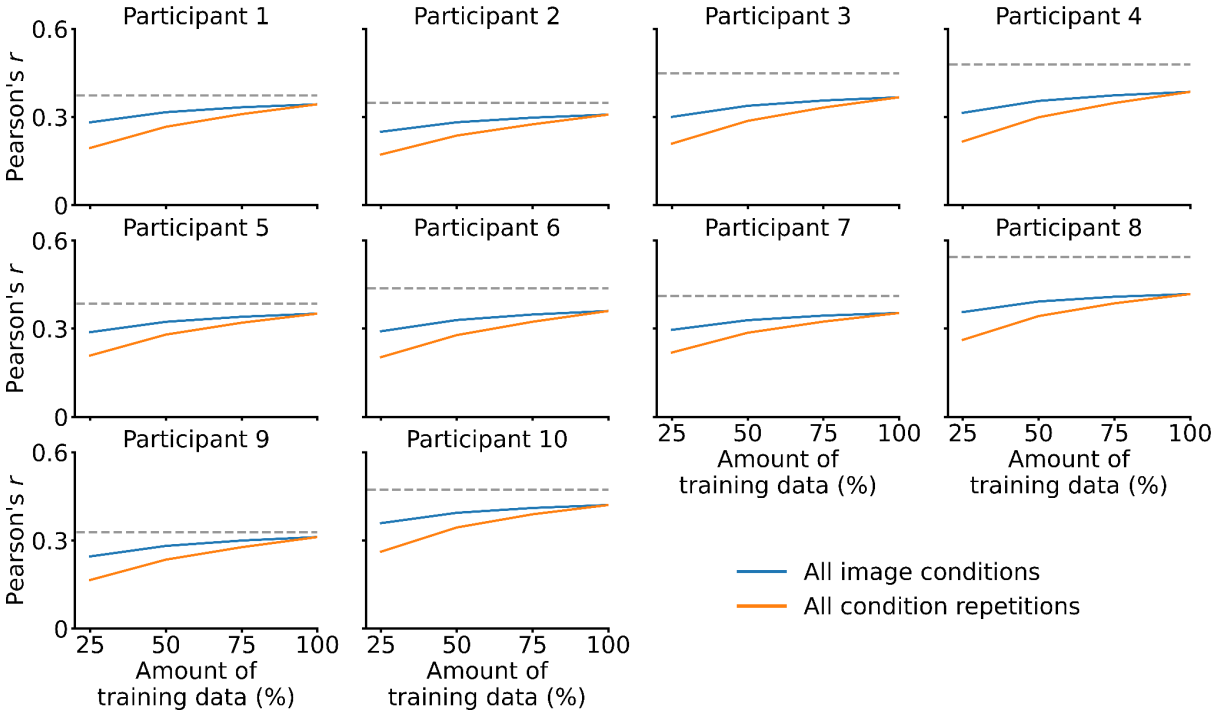
Supplementary Figure 13. For each SynTest ground truth image, we visualized the zero-shot identification algorithm best 10 candidate image conditions predictions (out of all 150,200 candidate images). Here we show the best 10 predictions of six SynTest image conditions for participant 1 (**A**), 6 (**B**) and 8 (**C**). Even when the correct image condition is not identified, our algorithm often selects image conditions conceptually and visually similar to the correct one.

Supplementary Figure 14: Effect of image conditions and condition repetitions, single participants



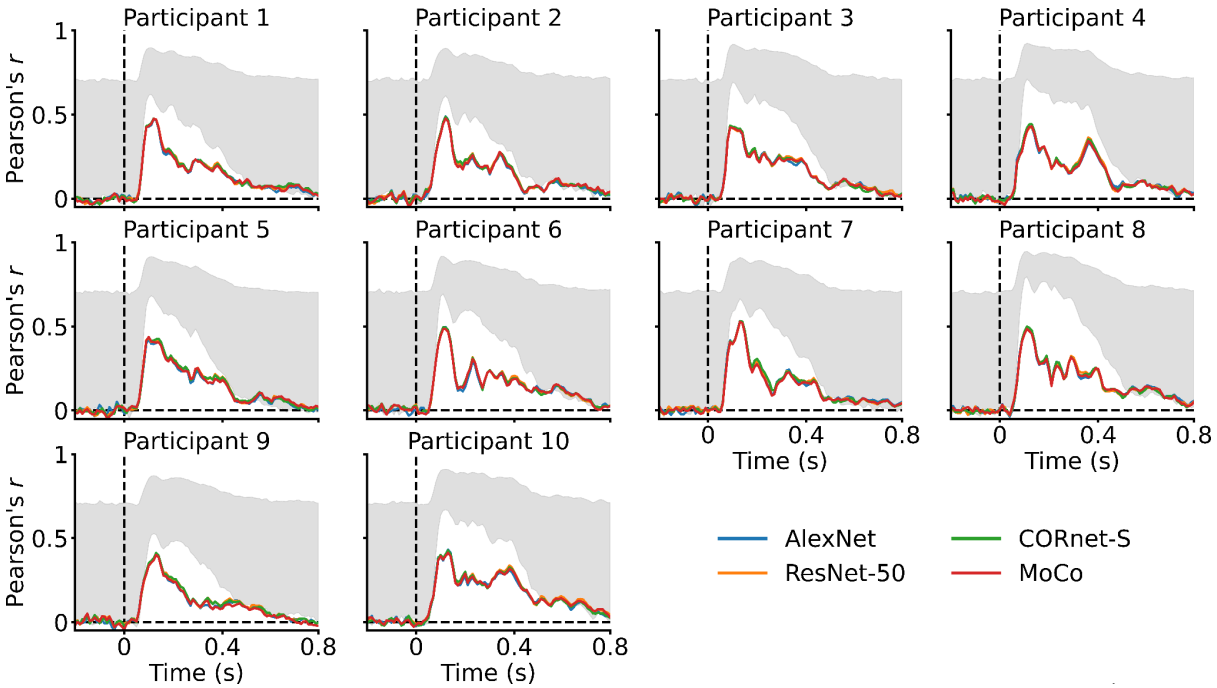
Supplementary Figure 14. Effect of image conditions and condition repetitions on linearizing encoding models' prediction accuracy, individual participants' results. Gray dashed lines as in **Figure 7**.

Supplementary Figure 15: Contribution of image conditions and condition repetitions, single participants



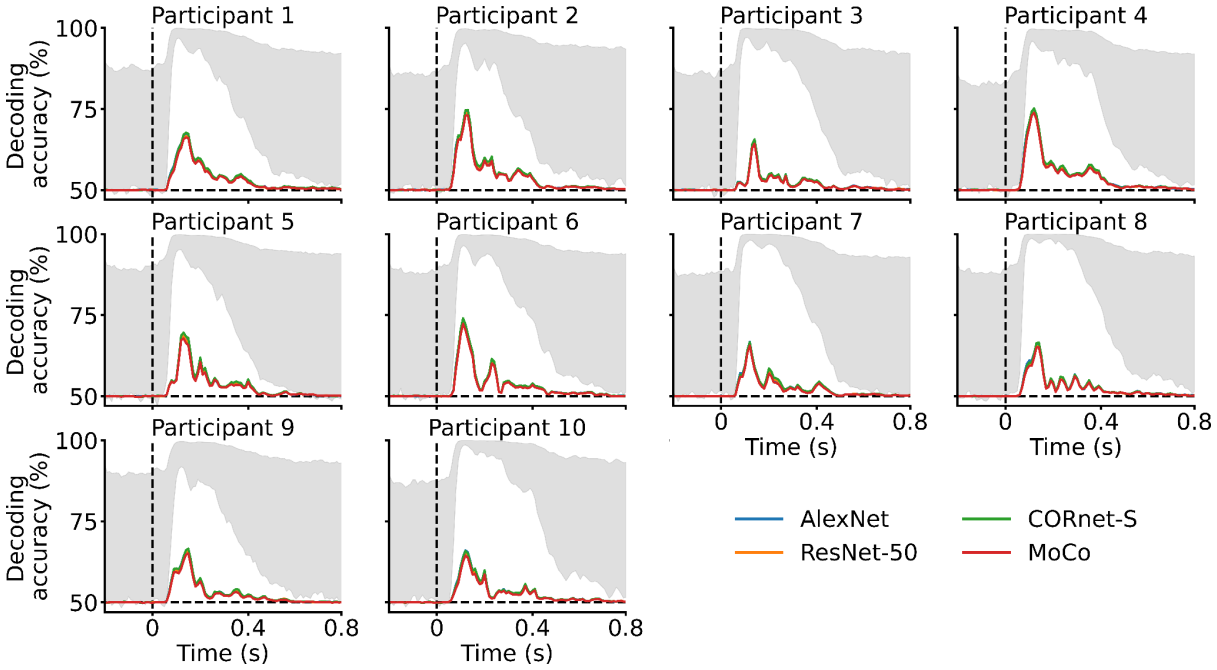
Supplementary Figure 15. Contribution of image conditions and condition repetitions on linearizing encoding models' prediction accuracy, individual participants' results. Gray dashed lines as in **Figure 7**.

Supplementary Figure 16: Inter-subject linearizing encoding, correlation results, single participants



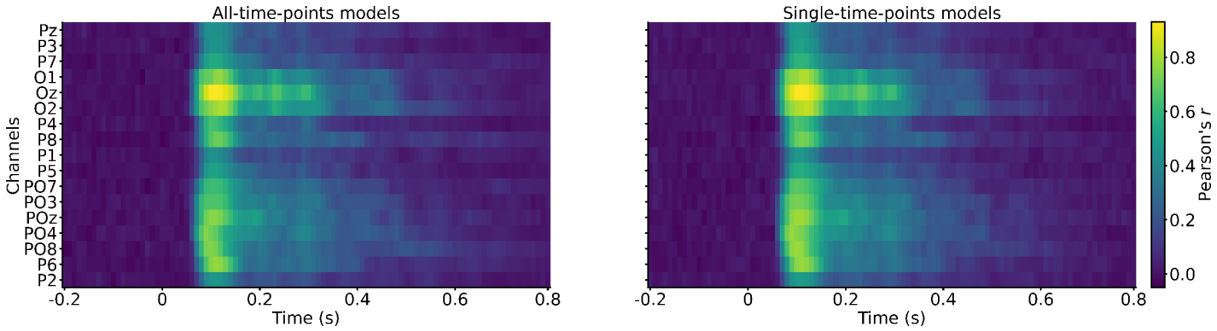
Supplementary Figure 16. Evaluating the prediction accuracy of linearizing encoding models which generalize to novel participants through correlation, individual participants' results. Gray areas and black dashed lines as in **Figure 4**.

Supplementary Figure 17: Inter-subject linearizing encoding, pairwise decoding results, single participants



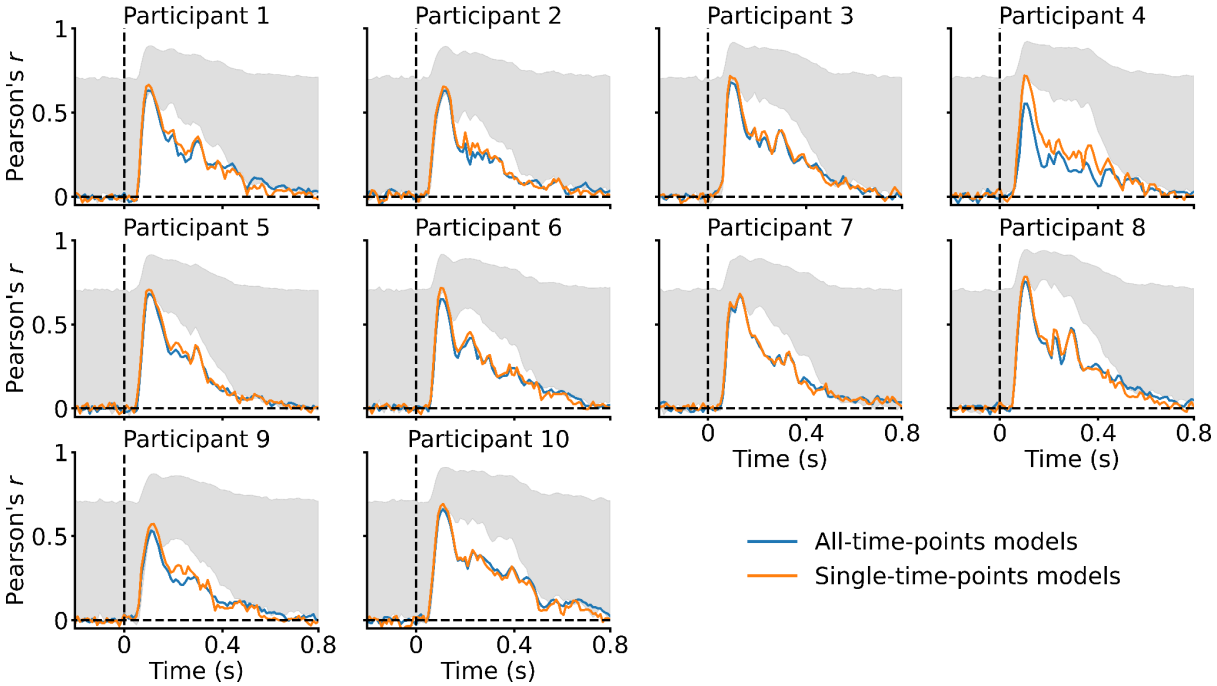
Supplementary Figure 17. Evaluating the prediction accuracy of linearizing encoding models which generalize to novel participants through pairwise decoding, individual participants' results. Gray areas and black dashed lines as in **Figure 4**.

Supplementary Figure 18: End-to-end encoding correlation results, single EEG channels



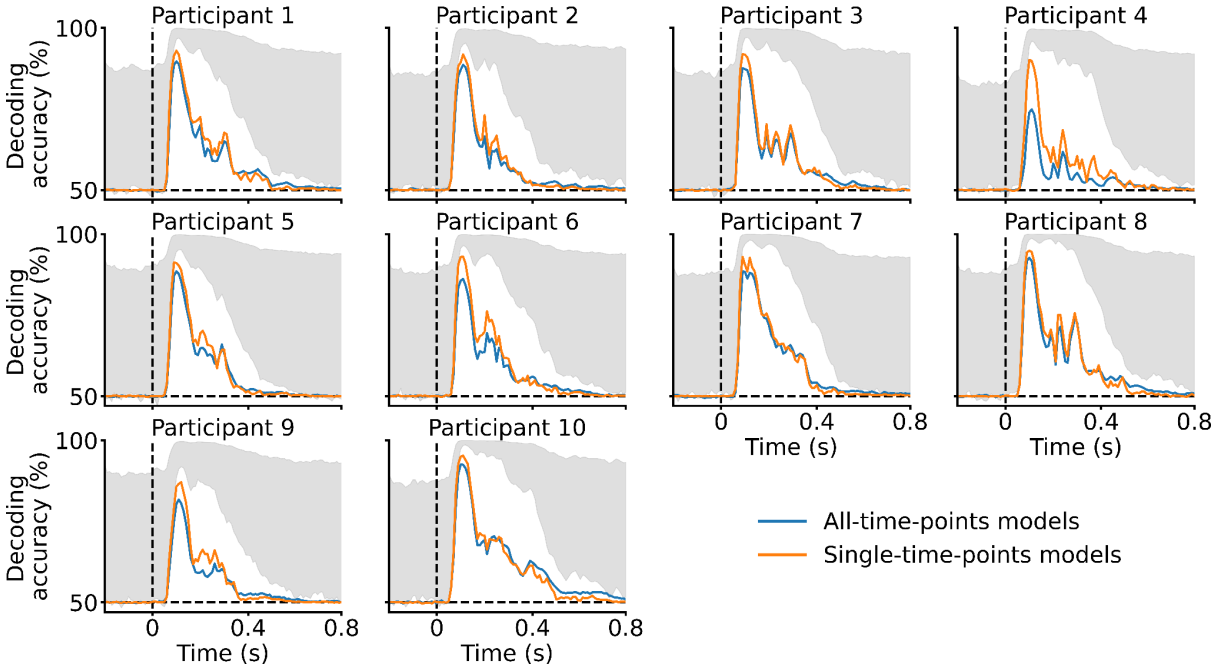
Supplementary Figure 18. Single EEG channels correlation results of the end-to-end encoding models, averaged across participants. Occipital channels are best modeled, followed by parieto-occipital channels and finally parietal channels.

Supplementary Figure 19: End-to-end encoding, correlation results, individual participants



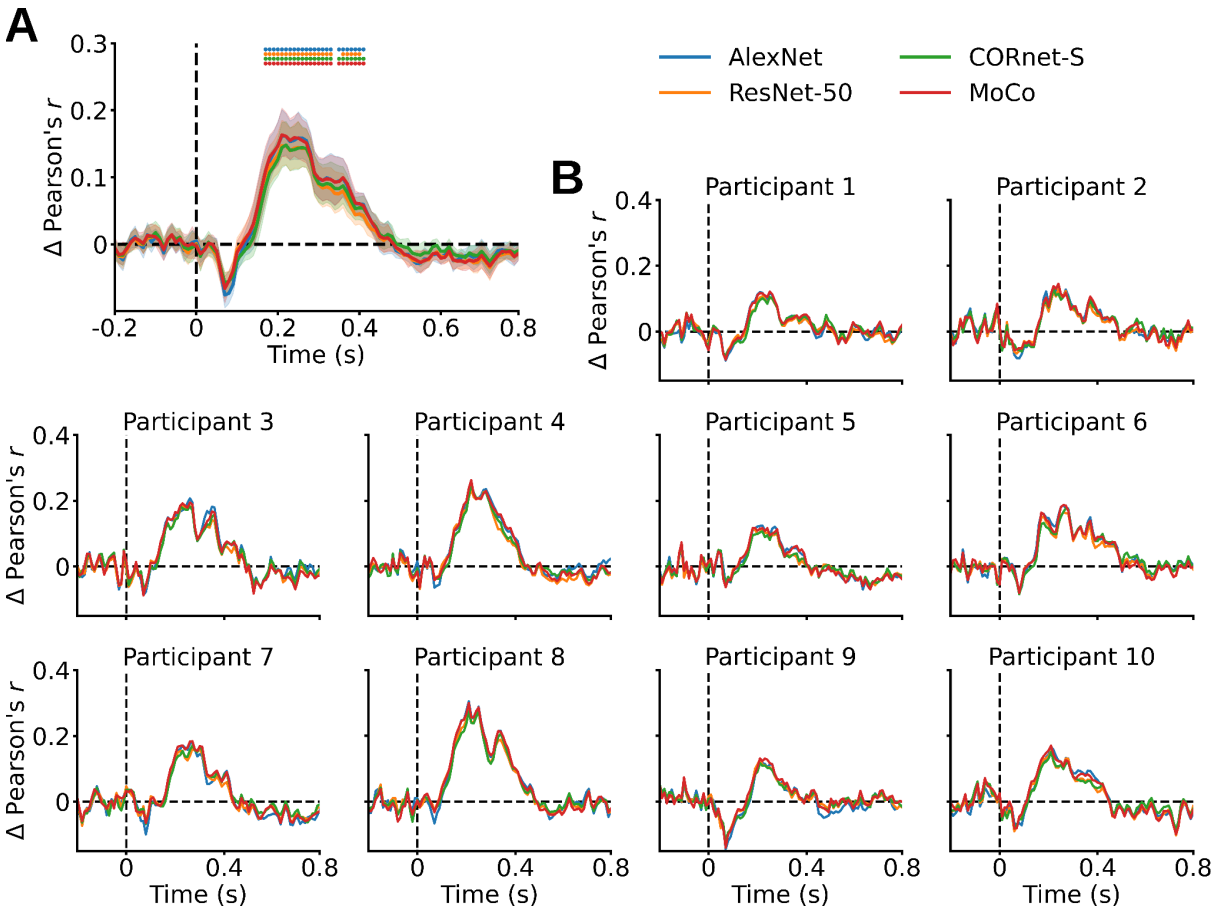
Supplementary Figure 19. Evaluating the end-to-end encoding models' prediction accuracy through correlation, individual participants' results. Gray areas and black dashed lines as in **Figure 4**.

Supplementary Figure 20: End-to-end encoding, pairwise decoding results, single participants



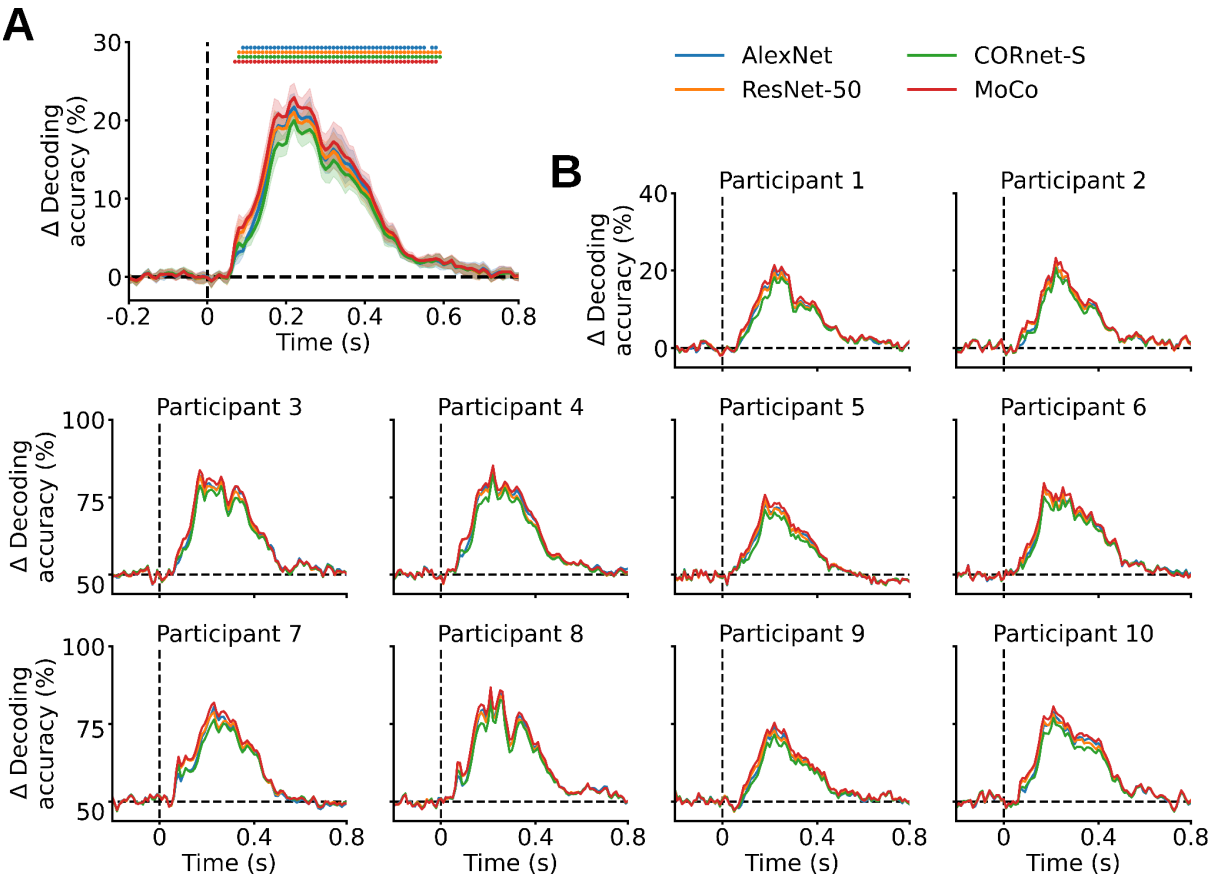
Supplementary Figure 20. Evaluating the end-to-end encoding models' prediction accuracy through pairwise decoding, individual participants' results. Gray areas and black dashed lines as in **Figure 4**.

Supplementary Figure 21: Differences between the correlation results and the noise ceiling lower bound estimates



Supplementary Figure 21. Difference between the correlation results and the noise ceiling lower bound estimates. (**A**) The differences averaged across participants are significantly above zero between 170ms and 410ms, with peaks at 210-220ms ($P < 0.05$, one-sided t-test after Fisher's z-transform, Bonferroni-corrected). (**B**) Individual participants' results. Error margins, asterisks and black dashed lines as in **Figure 4**.

Supplementary Figure 22: Differences between the pairwise decoding results and the noise ceiling lower bound estimates



Supplementary Figure 22. Difference between the pairwise decoding results and the noise ceiling lower bound estimates. (**A**) The differences averaged across participants are significantly above zero between 70ms and 590ms, with peaks at 220ms ($P < 0.05$, one-sample one-sided t-test, Bonferroni-corrected). (**B**) Individual participants' results. Error margins, asterisks and black dashed lines as in **Figure 4**.