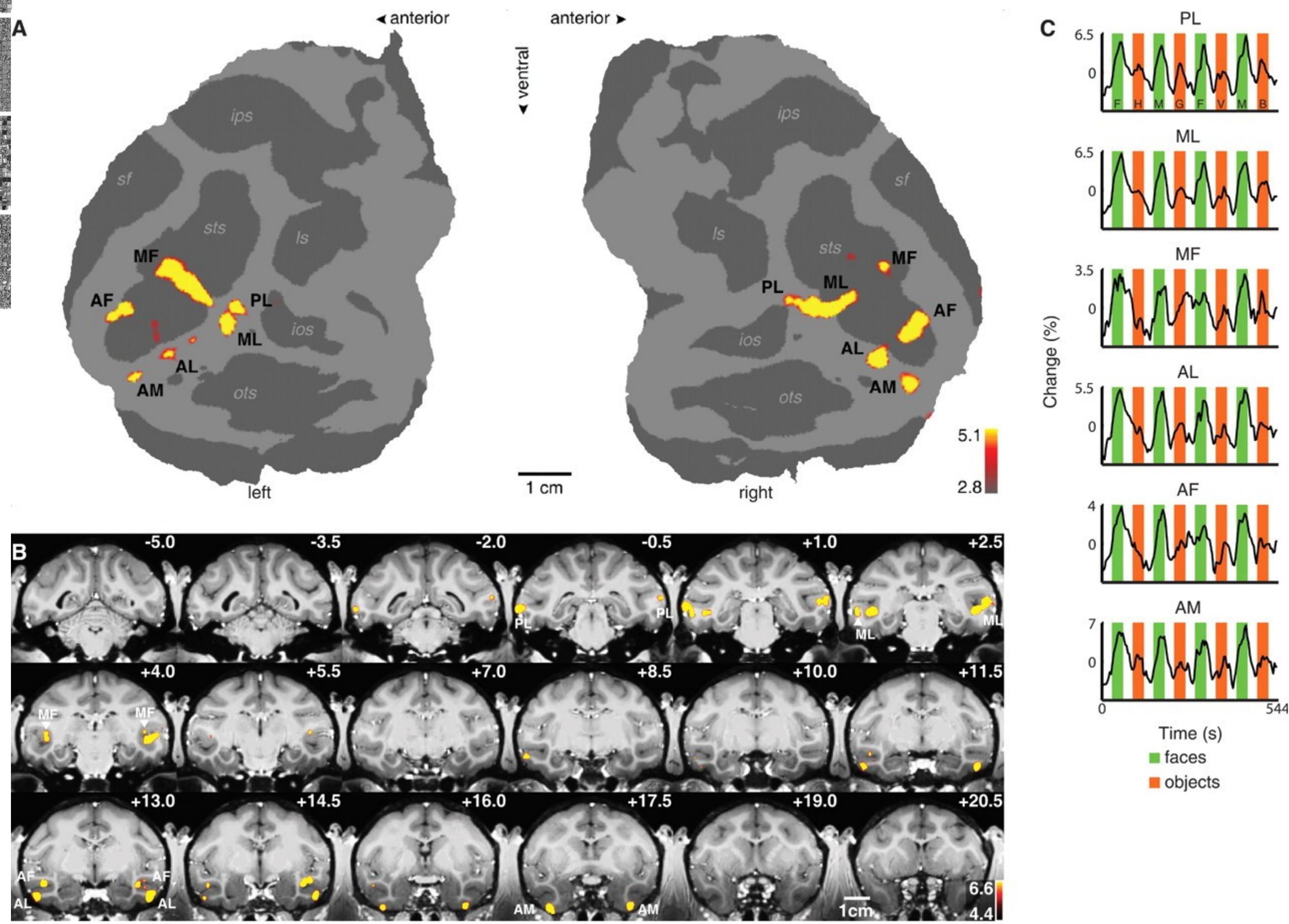
A specialized face-processing network consistent with the representational geometry of monkey face patches

Amirhossein Farzmahdi, Karim Rajaei, Masoud Ghodrati, Reza Ebrahimpour, Seyed-Mahdi Khaligh-Razavi

Grace Lindsay
3/4/15

Face Patches: Temporal cortex of Macaque shows several areas dedicated to face processing



"**six discrete face-selective regions**, consisting of one posterior face patch [posterior lateral (PL)], twomiddle face patches [middle lateral (ML) and middle fundus (MF)], and three anterior face patches [anterior fundus (AF), anterior lateral (AL), and anterior medial (AM)], spanning the entire extent of the temporal lobe(Moeller et al., 2008)...First in the hierarchy is PL, which contains a high concentration of face-selective cells, driven by the presence of **face components** (Issa and DiCarlo, 2012). Middle patches represent simple properties of faces (e.g. **face-views**) and in anterior parts, neurons become selective to more complex face properties (e.g.. **face identities**– Freiwald and Tsao, 2010)"

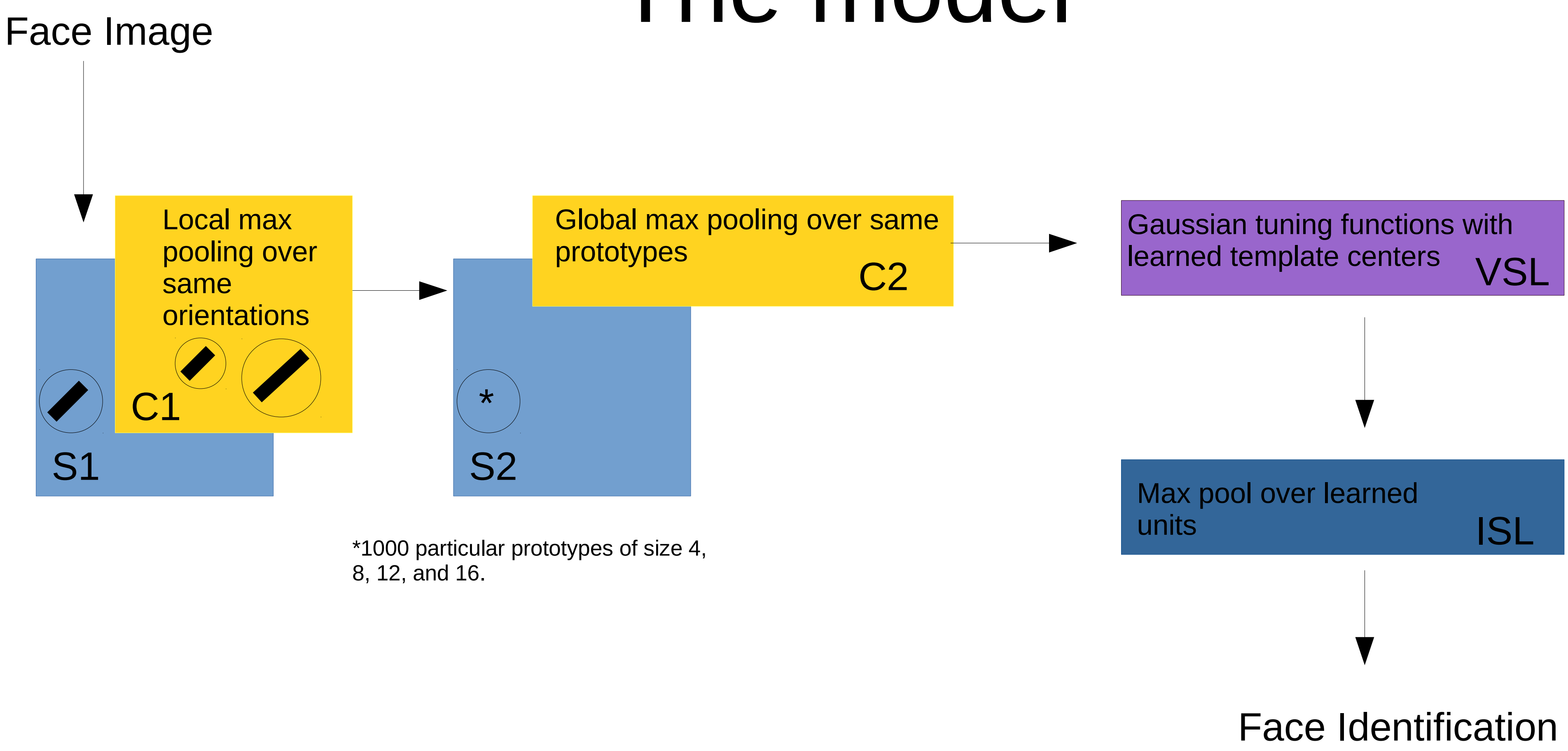
This Paper

Our proposed model of face processing is based on recent electrophysiological evidence in monkey face selective areas (Freiwald and Tsao, 2010; Moeller et al., 2008; Tsao et al., 2006).The model has several layers with an **organization similar to that of the hierarchical structure of the face processing system**. Layers of our model simulate different aspects of face processing and its representational space similar to that of monkey face patches (Freiwald and Tsao, 2010),. The model has view selective and identity selective layers **consistent with physiological and psychophysical data**.

6 layers:

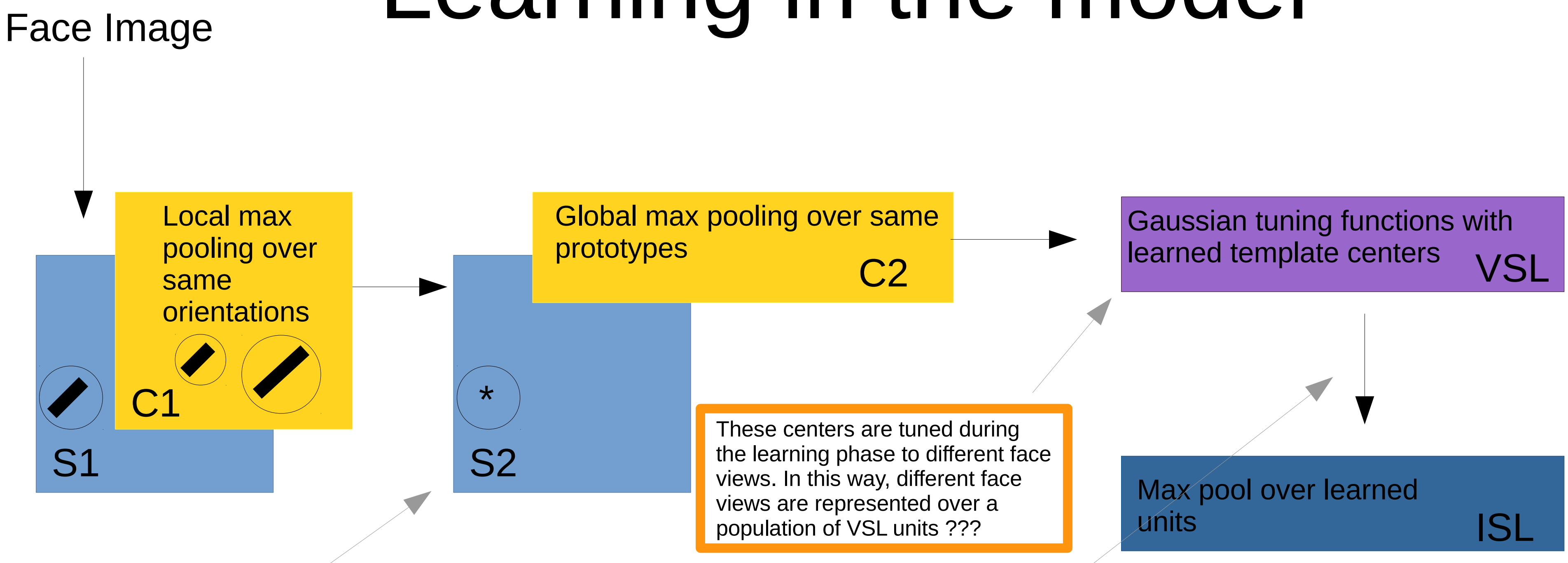
- First 4 layers = primary feature extraction (early cortex through PL)
- View selective layer (middle patches)
- Identity selective layer (anterior patches)

The model



*1000 particular prototypes of size 4, 8, 12, and 16.

Learning in the model



*1000 particular prototypes that are randomly extracted using an unsupervised random selection mechanism from training images.

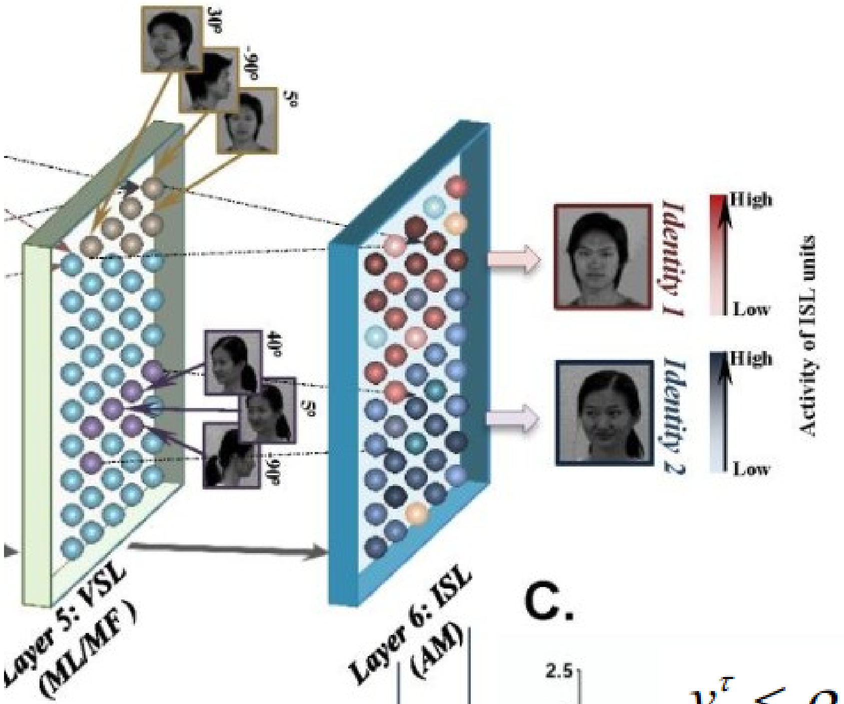
This is done by **correlating face views of the same identity across time** (temporal correlation); the idea being that in the real world, face views of an identity smoothly changes in time (abrupt changes of view are not expected). The time interval between face views of two identities (sequence of showing two identities)causes VSL units to make connections with different ISL units.

$$y_i^\tau = (1-\alpha)e^{-\frac{1}{2\sigma}(X-P_i)^2} + \alpha y_i^{\tau-1}$$

$$w_{ij} = \begin{cases} 1 & y_i^\tau > \rho \\ 0 & otherwise \end{cases}$$

$$z_j^\tau = (1-\beta).Max(y_i^\tau) + \beta.z_j^{\tau-1}$$

Creating view-invariant identification



$$y_i^\tau = (1-\alpha)e^{-\frac{1}{2\sigma}(X-P_i)^2} + \alpha y_i^{\tau-1}$$

$$w_{ij} = \begin{cases} 1 & y_i^\tau > \rho \\ 0 & otherwise \end{cases}$$

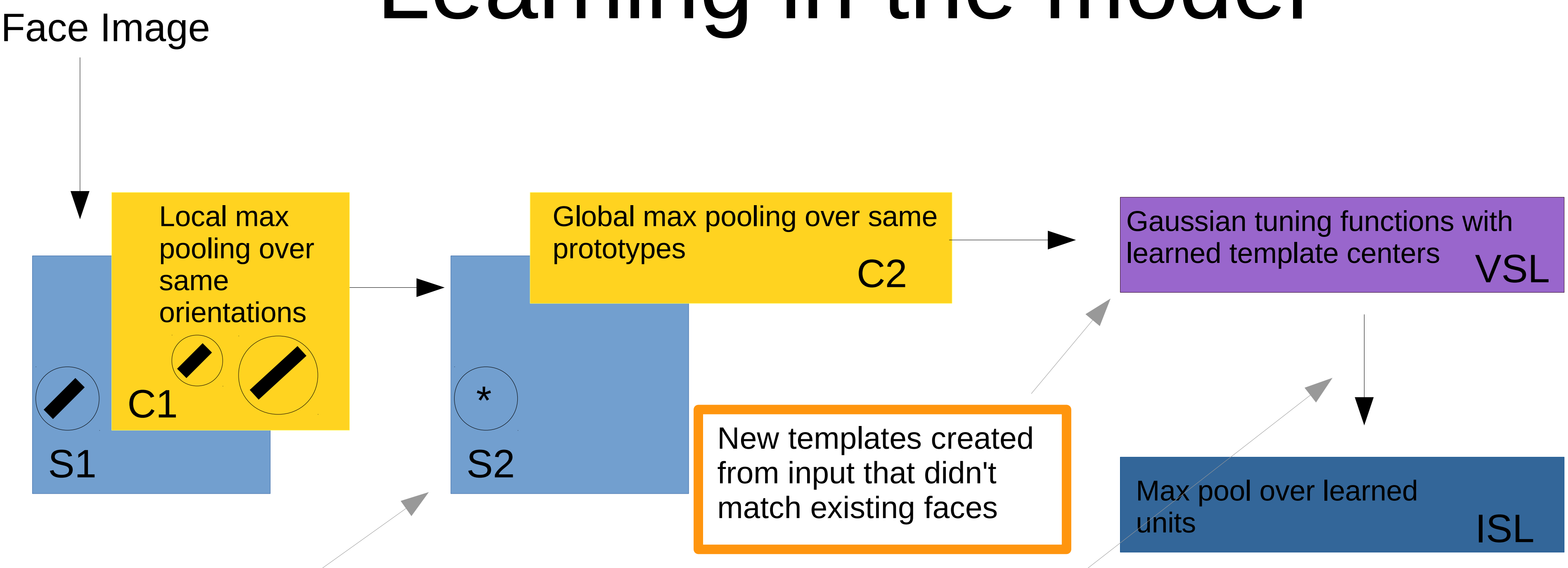
-50 identities with 37 views each. During training, images are presented in natural order and traces are wiped in between different identities

$y_i^\tau < \rho$ → a VSL unit with a Gaussian-like function that is tuned to the input stimuli is created

i=most recent added unit?

The discriminability between identities is measured and compared to the previous state of the model (before adding new units), using a View-invariant Identity Selectivity Index (VISI) and a support vector machine (SVM) classifier identification performance as measures of identity selectivity and invariant face recognition, respectively. The VISI value is compared with a threshold; **a value less than the threshold indicates that the new modification (units added to the model) had no significant impact on improving the discriminability. Therefore, the new added units are removed**

Learning in the model



*1000 particular prototypes that are randomly extracted using an unsupervised random selection mechanism from training images.

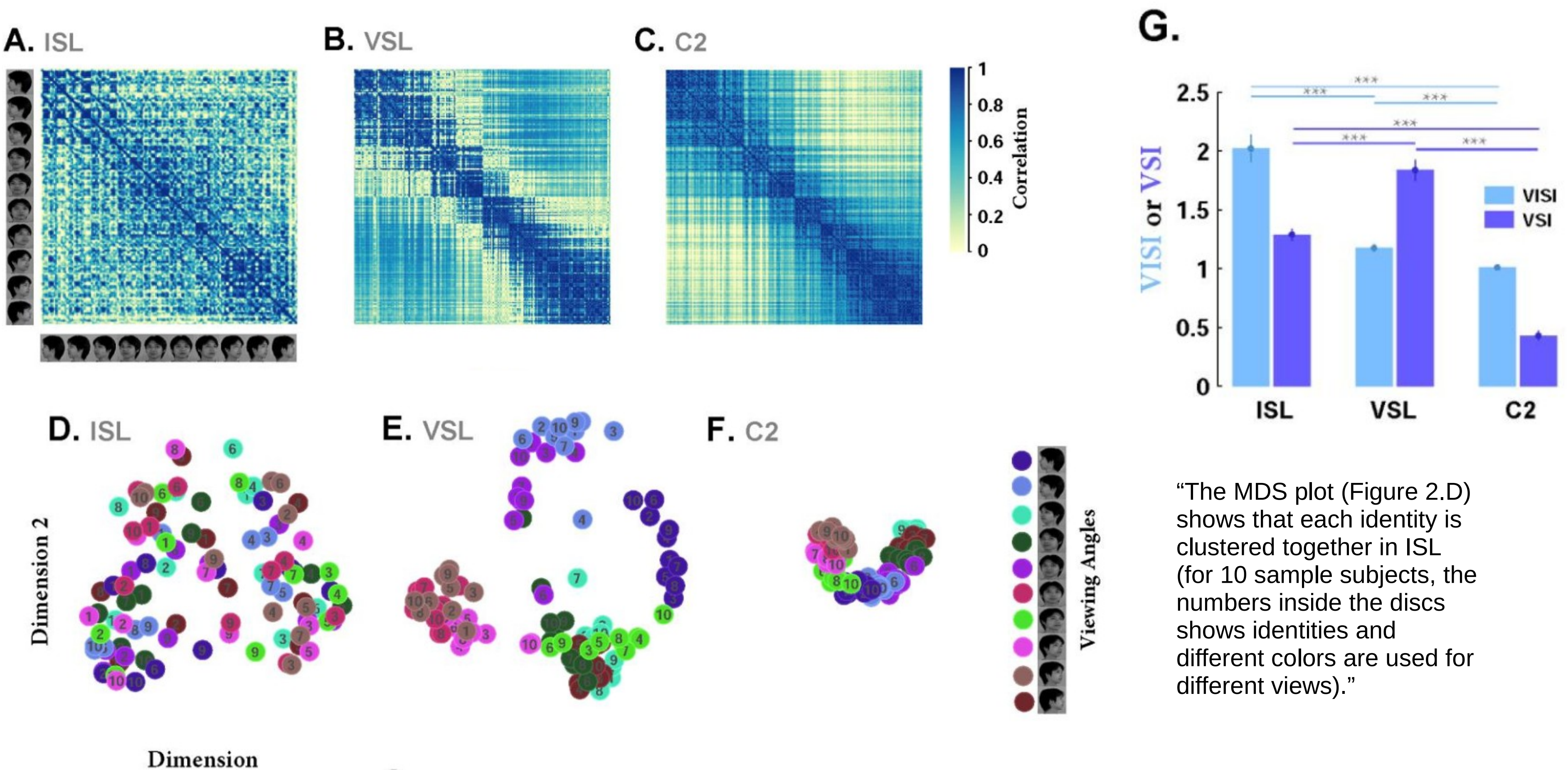
This is done by **correlating face views of the same identity across time** (temporal correlation); the idea being that in the real world, face views of an identity smoothly changes in time (abrupt changes of view are not expected). The time interval between face views of two identities (sequence of showing two identities)causes VSL units to make connections with different ISL units.

$$y_i^\tau = (1-\alpha)e^{-\frac{1}{2\sigma}(X-P_i)^2} + \alpha y_i^{\tau-1}$$

$$w_{ij} = \begin{cases} 1 & y_i^\tau > \rho \\ 0 & otherwise \end{cases}$$

$$z_j^\tau = (1-\beta).Max(y_i^\tau) + \beta.z_j^{\tau-1}$$

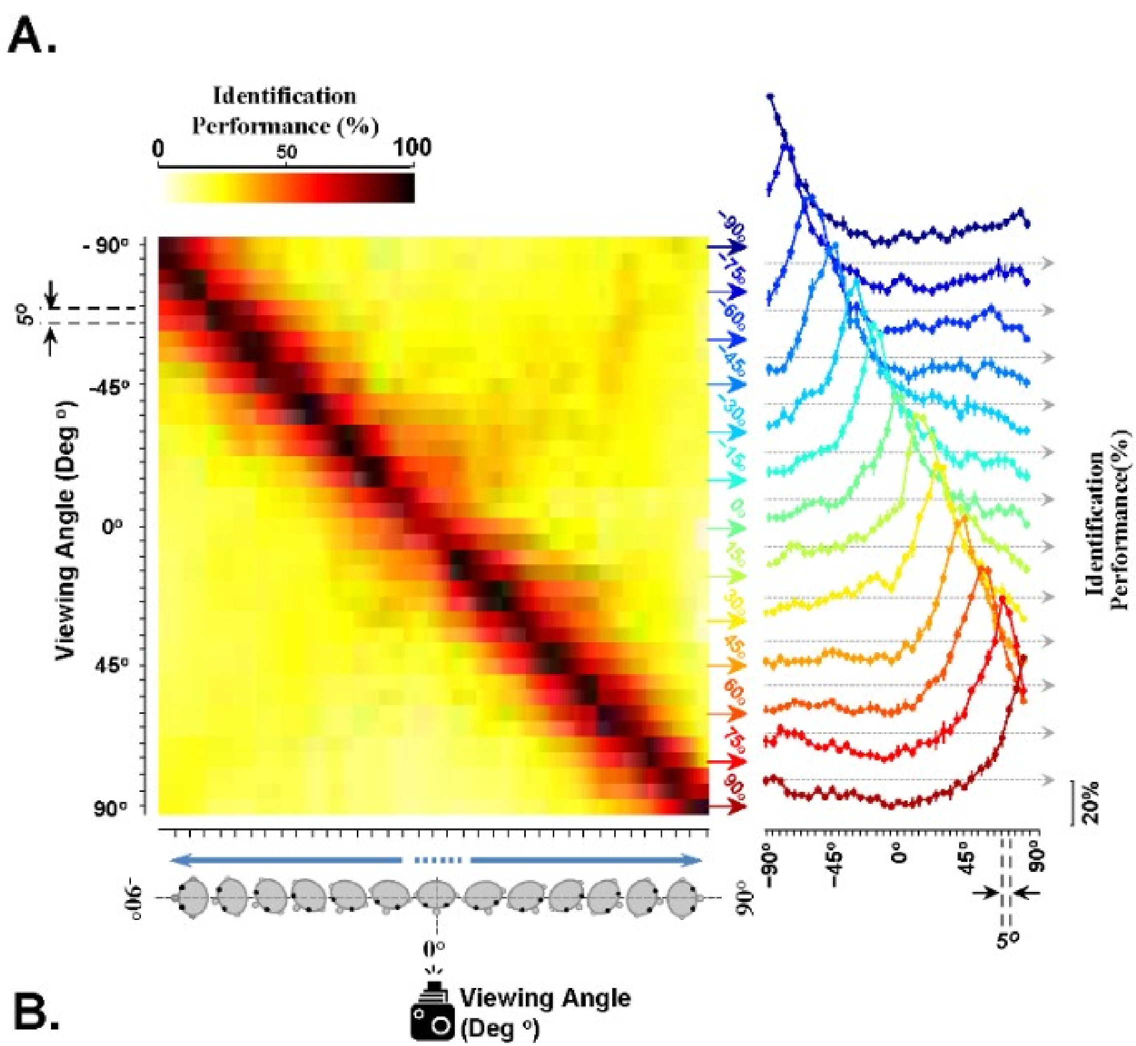
Model evaluation



"The MDS plot (Figure 2.D) shows that each identity is clustered together in ISL (for 10 sample subjects, the numbers inside the discs shows identities and different colors are used for different views)."

- View selectivity index (VSI) determines how correlated the representations of images with the same viewing angle are
- View-invariant identity selectivity index (VISI) determines how correlated the representations of images with the same identity are
- An SVM classifier is also trained on 18 face views of 20 identities of evaluation dataset (randomly selected from 37 face views) and tested on 19 face views

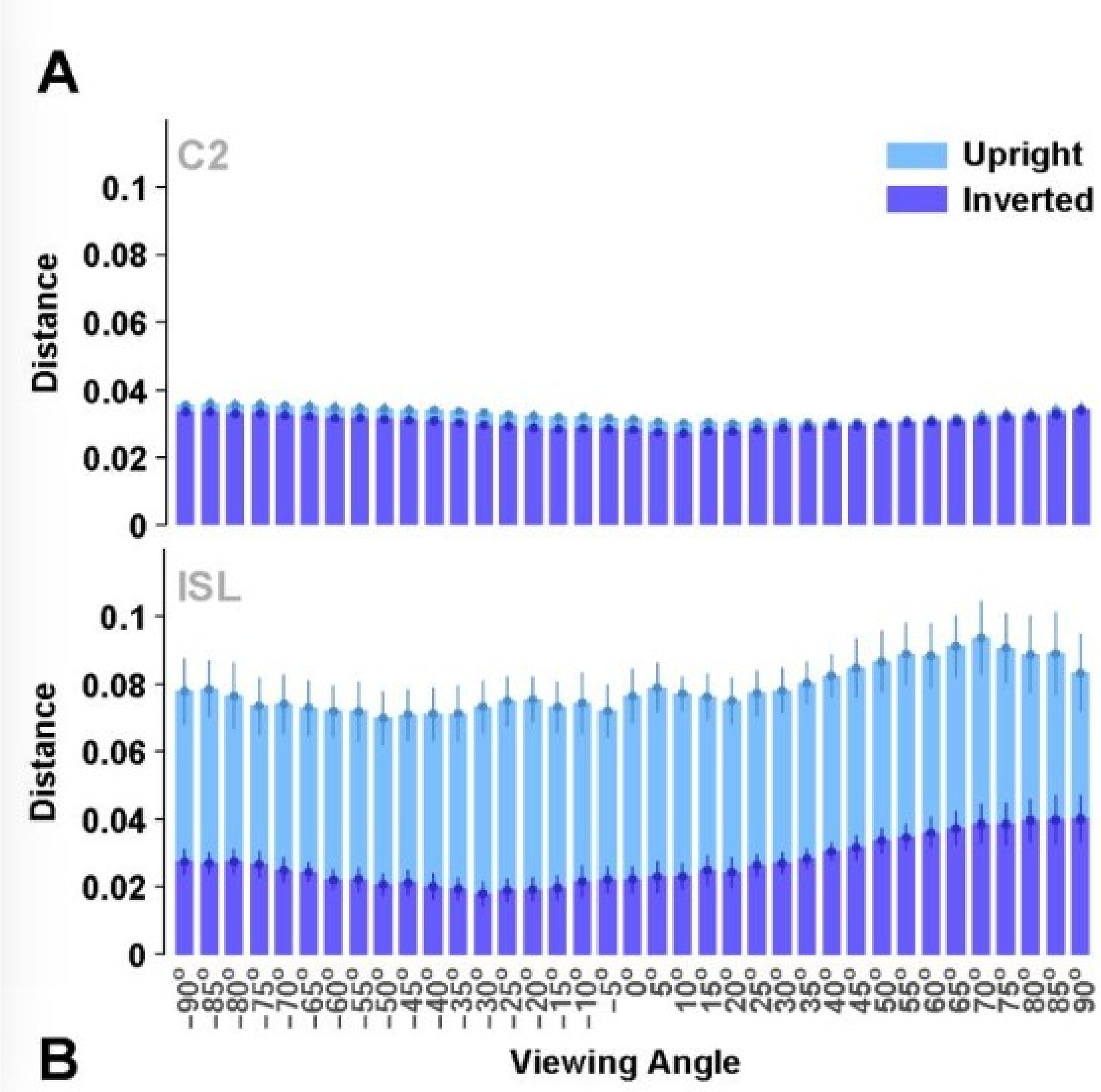
Canonical face view



The SVM was trained with one view and tested by other views (repeated across 10 individual runs for every view, separately). The performance decreases as the views deviate from the training view, Figure 4.A . This observation might not be surprising; but, the interesting point is that **the degree of invariance in ISL features increases around canonical face views**, Figure 4.B. These evaluations exhibit that the model is able to represent the effect of canonical face views

The idea of canonical face view refers to the observation that specific face views carry a higher amount of information about face identities, therefore face identification performance for these views is significantly higher

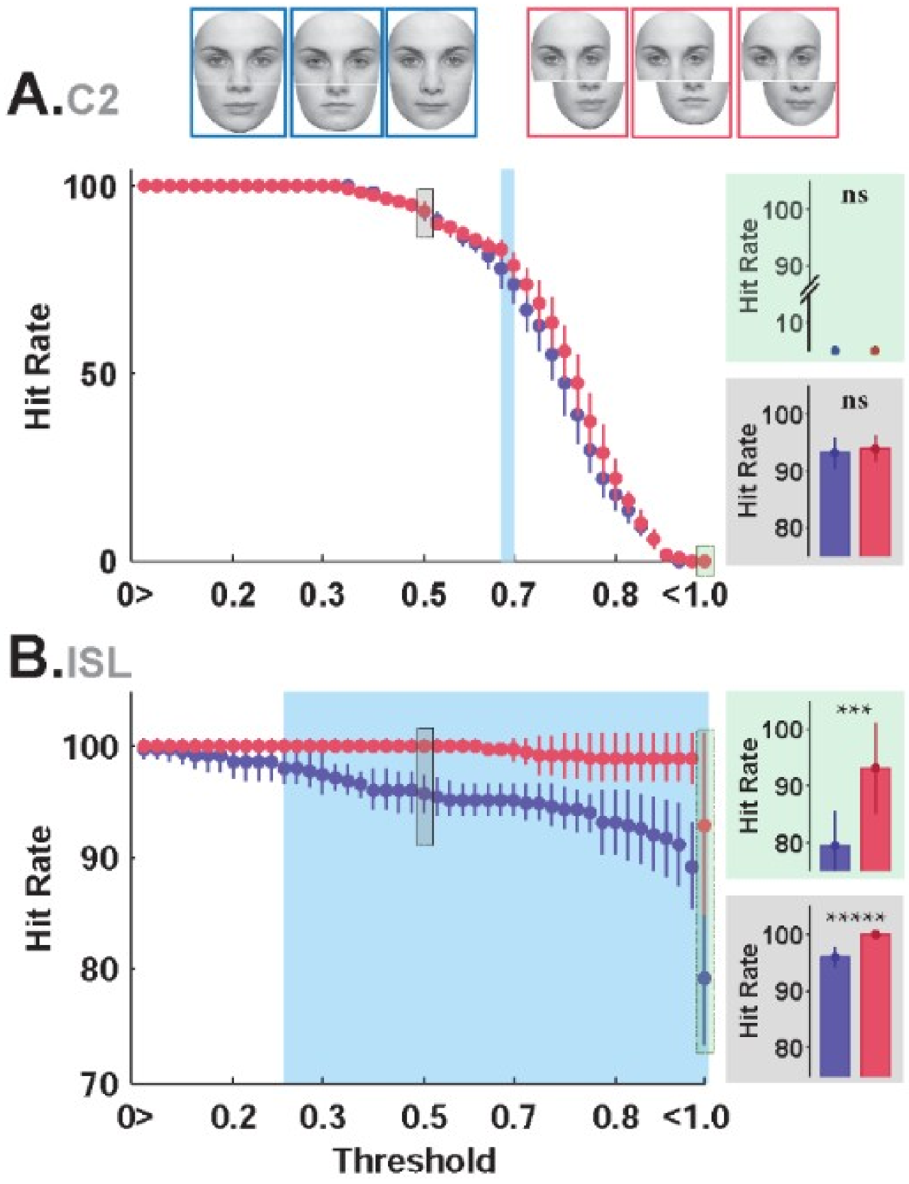
Inversion Test



The **distance between feature vectors of inverted and upright face images** for C2 units (up) and ISL (down). Inversion effect is highly significant at ISL compared to the C2 layer (normalized Euclidean distance). The vertical axis indicates the normalized distance and the horizontal axis shows different views, separated with the steps of 5º.

What is this distance?

Misalignment



For ISL, the hit rate in misaligned images (red curve) is significantly higher than the aligned faces (blue curve) for all thresholds above 0.25. This indicates that two identical top halves with misalignment are assumed more similar than the aligned case (i.e. having two identical top halves with aligned lower parts, which makes them to be perceived as different identities).There is no clear difference in C2 responses between aligned and misaligned faces, Figure 6.A.

To investigate CFE in the proposed model, we trained the model using NCKU dataset (Chen and Lien, 2009)-see Material and Methods for details). In the test phase, the model was presented with composite face stimuli from Rossion (2013),

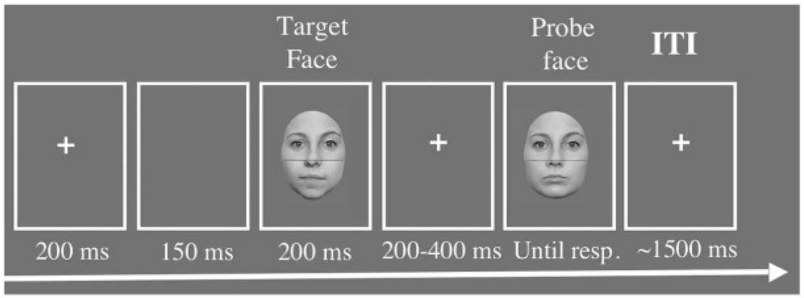
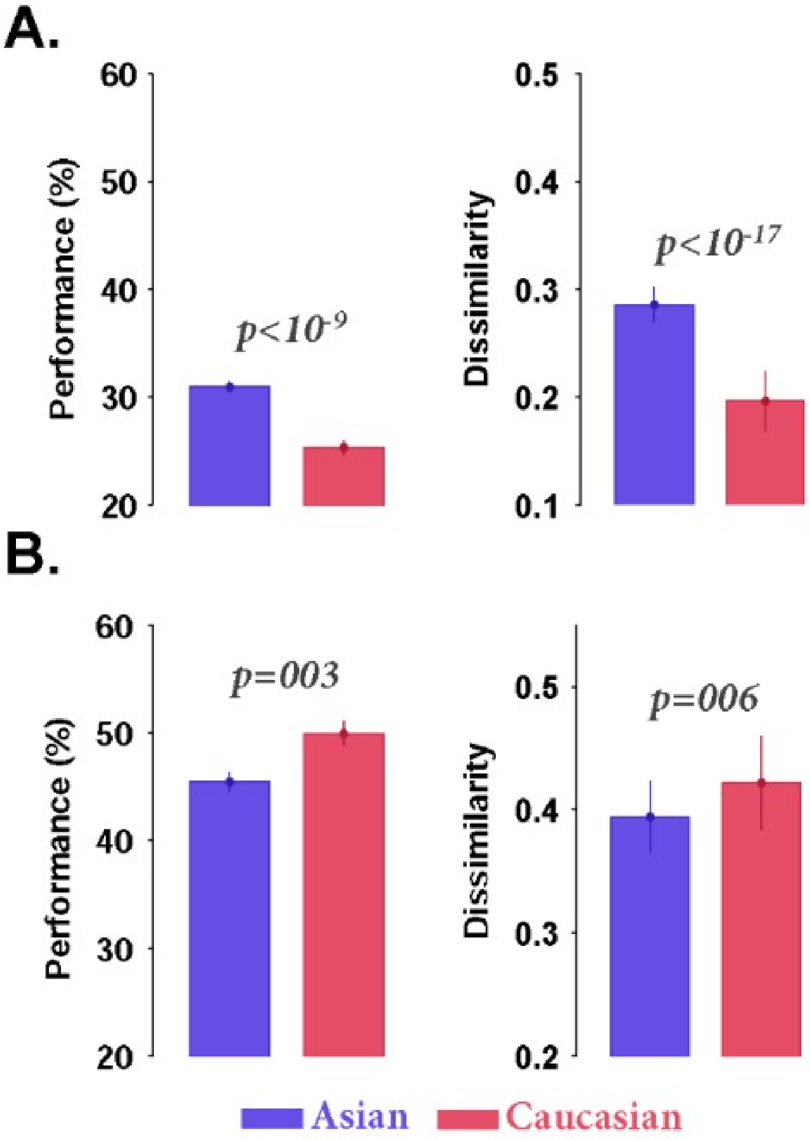


Figure 2. The composite face illusion in the context of a delayed matching task. Observers have to match the sequentially presented top halves (top = above the small gap between the face halves). The task is difficult because the top halves are erroneously perceived as being different.

Other Race Effect



The model was trained using images from NCKU dataset (Asian race)and tested using Asian and Caucasian images fromTarr dataset.

The model was trained using images from Tarr dataset (Caucasian race)and tested using Asian and Caucasian images from Tarr dataset.

Conclusions

- This model attempts to replicate properties of the face-processing system in macaques
- Seems likely that this model is over-specified and over-trained
- The analyses supporting their replication of certain psychophysical findings are flawed
- Good overview: “Mechanisms of face perception” by Doris Y. Tsao and Margaret S. Livingstone, 2008.