

2015

Normalization Among Heterogeneous Population Confers Stimulus Discriminability on the Macaque Face Patch Neurons

Akinori F. Ebihara

Follow this and additional works at: http://digitalcommons.rockefeller.edu/student_theses_and_dissertations



Part of the [Life Sciences Commons](#)

Recommended Citation

Ebihara, Akinori F., "Normalization Among Heterogeneous Population Confers Stimulus Discriminability on the Macaque Face Patch Neurons" (2015). *Student Theses and Dissertations*. 278.

http://digitalcommons.rockefeller.edu/student_theses_and_dissertations/278

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact mcsweej@mail.rockefeller.edu.



**NORMALIZATION AMONG HETEROGENEOUS
POPULATION CONFERS STIMULUS
DISCRIMINABILITY ON THE MACAQUE FACE
PATCH NEURONS**

**A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy**

by

Akinori F. Ebihara

June 2015

**NORMALIZATION AMONG HETEROGENEOUS
POPULATION CONFERS STIMULUS DISCRIMINABILITY
ON THE MACAQUE FACE PATCH NEURONS**

Akinori F. Ebihara, Ph.D.

The Rockefeller University 2015

Primates are capable of recognizing faces even in highly cluttered natural scenes. In order to understand how the primate brain achieves face recognition despite this clutter, it is crucial to study the representation of multiple faces in face selective cortex. However, contrary to the essence of natural scenes, most experiments on face recognition literatures use only few faces at a time on a homogeneous background to study neural response properties. It thus remains unclear how face selective neurons respond to multiple stimuli, some of which might be encompassed by their receptive fields (RFs), others not. How is the neural representation of a face affected by the concurrent presence of other stimuli? Two lines of evidence lead to opposite predictions: first, given the importance of MAX-like operations for achieving selectivity and invariance, as suggested by feedforward circuitry for object recognition, face representations may not be compromised in the presence of clutter. On the other hand, the psychophysical crowding effect - the reduced discriminability (but not detectability) of an object in clutter - suggests that an object representation may be impaired by additional stimuli. To address this question, we conducted electrophysiological recordings in the macaque temporal lobe, where bilateral face selective areas are tightly interconnected to form a hierarchical face processing stream. Assisted by functional MRI, these face patches could be targeted for single-cell recordings. For each neuron, the most preferred face stimulus was determined, then presented at the center of the neuron's

RF. In addition, multiple stimuli (preferred or non-preferred) were presented in different numbers (0,1,2,4 or 8), from different categories (face or non-face object), or at different proximity (adjacent to or separated from the center stimulus). We found the majority of neurons reduced mean firing rates more (1) with increasing numbers of distractors, (2) with face distractors rather than with non-face object distractors, (3) at closer distractor proximity, and, additionally, (4) the response to multiple preferred faces depends on RF size. Although these findings in single neurons could indicate reduced discriminability, we found that each stimulus condition was well separated and decodable in a high-dimensional space spanned by the neural population. We showed that this was because neuronal population was quite heterogeneous, yet changing response systematically as stimulus parameter changed. Few neurons showed MAX-like behavior. These findings were explained by divisive normalization model, highlighting the importance of the modular structure of the primate temporal lobe. Taken together, these data and modeling results indicate that neurons in the face patches acquire stimulus discriminability by virtue of the modularity of cortical organization, heterogeneity within the population, and systematicity of the neural response.

Acknowledgements

First and foremost, I would like to thank my adviser Winrich Freiwald, for his invaluable guidance, patience and encouragement. It was fortunate for me to matriculate Rockefeller University same time as Winrich's inauguration - as his very first student in Rockefeller, I could have plenty of time discussing variety of topics including my thesis project, reckless experiment and lab setups. Building a lab from scratch entailed lots of difficulties, but it will definitely help me in some future.

I also thank Marcelo Magnasco, my co-advisor. His advice was always insightful and provided a different way of thinking on the same subject. With my collaborator, Ana Hocevar, he supported the modeling part and analysis of the project. Ana and I had numerous meetings to discuss model structure, data evaluation and interpretation. Without their assistance, this project could not be finalized. I sincerely thank them.

Fantastic committee members, Cori Bargmann and Charles Gilbert, they always provided sharp questions and suggestions during the Faculty Advisory Committee meetings to prevent the project from deflection. Throughout the discussion, I could feel myself in transition from just an ambitious student to a matured scientist with critical thinking. I was truly lucky to be able to spend time with them.

I am grateful to my parents and friends, who patiently supported my long, long journey to the Ph.D. Whenever I talk to them they would listen with curiosity and gave me warm encouragements. And, of course, I sincerely thank my wife, Yuka. Every single chat with her was soothing yet intellectually stimulating.

I also thank all of the current and former Freiwald lab members, espe-

cially Lucy Petro and Wilbert Zarco for the experimental support, and Michael Borisov for programming help. They were also great people to hang out with.

All the people in Dean's office - Sid Strickland, Emily Harms, Cris Rosario, Kristen Cullen, Marta Delgado, they're always available and supportive. I am grateful about that.

Overall, the time in the graduate school was truly enjoyable, productive and stimulating. I need to thank so many people all of whom cannot be listed here, for letting me spend exciting and fruitful years.

Table of Contents

List of Figures	ix
List of Tables	xiii
List of Equations	xvii
1 Introduction	1
1.1 Faces as invaluable sources of social information	1
1.2 Hierarchical organization of the visual cortex	4
1.3 Neurophysiological evidence of face selective neurons in the tem- poral lobe	6
1.4 Face selective neurons are segregated into modular structures . .	9
1.5 Early evidence of multiple stimulus representations and pro- posed models	10
1.6 Evidence of multiple stimulus representations in IT cortex . . .	13
1.7 Multiple stimuli representations in the macaque temporal lobe face patch	16
1.8 Organization of the thesis	17
2 Experimental design	19
2.1 Varying distractor numbers to study neural computation of mul- tiple visual inputs	19

2.2	Varying distractor categories in light of feature / conjunction search	20
2.3	Varying distractor proximity / eccentricity in light of the crowding effect	26
2.4	Stimulus selection and composition of multiple stimuli	29
2.5	Targeting the face patch MF and AL	39
3	Single neuron response properties and population average	41
3.1	Single neuron response and population average to Category/Number variation experiment	41
3.2	Single neuron response and population average to Proximity/Eccentricity variation experiment	50
3.3	Response dependency on receptive field size	55
3.4	Response reduction was not dependent on preferred face position	58
3.5	Summary	60
4	Population readout	62
4.1	Representation in the pattern of neural responses	62
4.2	How to capture representation in the high dimensional space spanned by population	65
4.3	Independent representations of the multiple stimuli in the population activity pattern	67
4.4	Quantifying discriminability assessing temporal transition of representation	72
5	Modeling results	76
5.1	Divisive Normalization framework to explain discriminability . .	76
5.2	Divisive normalization model explained the early phase of Category/Number variation experiment without any free parameter	77

5.3	Limitation of the pure prediction model and introducing an exponent as a free parameter	81
5.4	Divisive Normalization model revealed the change of computational principle from the early to the late phase	83
6	Comparison to anterior face patch AL	93
6.1	Anterior face patch AL is located one position higher than MF in the cortical hierarchy	93
6.2	AL neurons' responses were reduced by distractors similar to MF, but latency was shifted with distractor numbers	97
6.3	AL neural population preserved stimulus information in the representation space, revealed by ICA	103
6.4	Divisive Normalization explained the responses of AL neurons .	105
6.5	Summary	108
7	Discussion	110
7.1	Heterogeneity and systematicity with Divisive Normalization represent multiple stimuli in population activity	110
7.2	Modular structure confers category discriminability	113
7.3	Effect of modularity, heterogeneity and systematicity on representation space	115
7.4	Does a preferred face “pop-out”, or is it “crowded” by distractors?	121
7.5	Pop-out effect and attentional confound	122
7.6	Divisive Normalization gives an unified account encompassing previous seemingly contradictory results	124
7.7	Divisive Normalization to prevent saturation and to have economical representation	125

7.8	Representation at the late phase: robust against distractors yet maintaining representation	126
7.9	Transition from MF to AL	127
7.10	Concluding remark	128
A	Experimental Procedures	130
B	Divisive normalization model	137
C	Converting anatomical plane coordinate to stereotaxic arm coordinate	140
D	Converting pixels to degrees of visual angle	144
E	Finding the screen distance equalizing visual angle and im- age size	147
F	Population average without latency subtraction	149
G	Population average firing rate to multiple preferred faces.	151
H	Visualizing the population average in 2D space	153
I	Relationship between PCA and SVD	155
J	Tukey's Post-hoc test results	158
	Bibliography	177

List of Figures

1.1	Faces as information sources.	2
1.2	Human excels at face recognition.	3
1.3	Face selective neurons found in macaque temporal lobe.	7
1.4	Face selective region in human temporal lobe revealed by fMRI.	9
2.1	Distractor number variation.	21
2.2	Example of Feature search.	23
2.3	Example of Conjunction search.	24
2.4	Distractor category variation.	25
2.5	Psychophysical crowding effect.	27
2.6	Distractor proximity variation.	30
2.7	The target face selective cortex, MF.	31
2.8	Selection of stimuli and stimulus design.	34
2.9	Design of multiple stimuli.	36
2.10	Experiment flow chart.	38
3.1	Example neuronal response to the Cat/Num stimuli.	43
3.2	Population average response to the Cat/Num stimuli.	46
3.3	Significance test from ANOVA and Tukey's post-hoc test.	48
3.4	Coefficient from rational function fitting to compare distractor impact.	51
3.5	An example neural response to the Prox/Ecc experiment.	54

3.6	Population average response to the Prox/Ecc experiment.	56
3.7	Response reduction dependency on RF size.	57
3.8	Position control experiment.	61
4.1	Comparison of average, PCA and ICA.	64
4.2	PCA applied to the Cat/Num experiment.	68
4.3	DPCA applied to the Cat/Num experiment.	69
4.4	The concept of dimension reduction followed by ICA	70
4.5	High-dimensional representation preserves stimulus discriminabil- ity.	73
4.6	Decoding analysis on population activity.	75
5.1	The concept of Divisive Normalization.	78
5.2	Pure prediction model showed close correspondence to the in- dependently measured data.	80
5.3	Pure prediction model showed a larger error for the Prox/Ecc experiment.	82
5.4	Post-diction model explains the observed representation.	84
5.5	Robustness index calculated for the Cat/Num experiment.	86
5.6	Robustness index calculated for the Prox/Ecc experiment	87
5.7	Simulation of homogeneous and heterogeneous neural population.	90
5.8	Effect of modularity and systematicity.	92
6.1	Face patch connectivity diagram.	94
6.2	Comparison of physiological property between MF and AL neu- rons.	96
6.3	Response onset latency comparison between MF and AL.	98
6.4	AL population average firing rate to Cat/Num experiment.	100

6.5	Response onset-to-peak time, compared between AL and MF neurons.	102
6.6	AL population average firing rate to Prox/Ecc experiment. . . .	104
6.7	ICA on AL experimental results.	106
6.8	Pure prediction model on AL distractor category/number vari- ation experiment.	107
6.9	Postdiction model on AL experiment and population readout. .	109
7.1	Systematic and heterogeneous responses discriminate stimulus groups.	111
7.2	Modular structure confers category-discriminability.	114
7.3	Schematic drawing showing how ICA finds independent repre- sentations from dimension-reduced data.	116
7.4	PCA and ICA on a complete homogeneous response matrix. . .	117
7.5	PCA and ICA on a systematic and homogeneous response matrix.	117
7.6	PCA and ICA on a systematic, modular and homogeneous re- sponse matrix.	118
7.7	PCA and ICA on a systematic and heterogeneous response matrix.	119
7.8	PCA and ICA on a systematic, modular and heterogeneous re- sponse matrix.	120
C.1	Converting anatomical plane coordinates into stereotaxic arm coordinates.	141
D.1	Converting pixels to degrees.	145
E.1	Relating viewing distance to image size on the screen.	148
F.1	Population average without latency subtraction.	150

G.1	Population average firing rate to multiple preferred faces.	152
H.1	Projection to diagonal.	153

List of Tables

3.1	Cat/Num experiment, MF single cell average response	44
3.2	Cat/Num experiment, MF population average response	45
3.3	Cat/Num distractor-only control, MF population average response	49
3.4	Prox/Ecc experiment, MF single cell average response	52
3.5	Prox/Ecc experiment, MF population average response	53
3.6	Cat/Num experiment, average of 15 representative neurons . . .	59
3.7	Position control, average of 15 representative neurons	59
3.8	Cat/Num distractor-only control, average of 15 representative neurons	59
3.9	Position control distractor-only, average of 15 representative neurons	60
5.1	Data and the model average response and SEM	81
6.1	Cat/Num experiment, AL population average response	99
6.2	Prox/Ecc experiment, AL population average response	103
7.1	Role of modularity, heterogeneity and systematicity.	121
J.1	MF single cell example of Cat/Num experiment, face distractor conditions at the early phase	158
J.2	MF single cell example of Cat/Num experiment, face distractor conditions at the late phase	159

J.3	MF single cell example of Cat/Num experiment, object distractor conditions at the early phase	159
J.4	MF single cell example of Cat/Num experiment, object distractor conditions at the late phase	160
J.5	MF population average of Cat/Num experiment, face distractor conditions at the early phase	160
J.6	MF population average of Cat/Num experiment, face distractor conditions at the late phase	161
J.7	MF population average of Cat/Num experiment, object distractor conditions at the early phase	161
J.8	MF population average of Cat/Num experiment, object distractor conditions at the late phase	162
J.9	MF single cell example of Prox/Ecc experiment, RF center at the early phase	162
J.10	MF single cell example of Prox/Ecc experiment, RF center at the late phase	163
J.11	MF single cell example of Prox/Ecc experiment, RF periphery at the early phase	163
J.12	MF single cell example of Prox/Ecc experiment, RF periphery at the late phase	163
J.13	MF population average of Prox/Ecc experiment, RF center at the early phase	164
J.14	MF population average of Prox/Ecc experiment, RF center at the late phase	164
J.15	MF population average of Prox/Ecc experiment, RF periphery at the early phase	165

J.16	MF population average of Prox/Ecc experiment, RF periphery at the late phase	165
J.17	MF population average of Prox/Ecc experiment, distractor only control, RF center at the early phase	166
J.18	MF population average of Prox/Ecc experiment, distractor only control, RF center at the early phase	166
J.19	MF population average of Prox/Ecc experiment, distractor only control, RF periphery at the early phase	167
J.20	MF population average of Prox/Ecc experiment, distractor only control, RF periphery at the early phase	167
J.21	AL population average of Cat/Num experiment, face distractor conditions at the early phase	168
J.22	AL population average of Cat/Num experiment, face distractor conditions at the late phase	169
J.23	AL population average of Cat/Num experiment, object distrac- tor conditions at the early phase	169
J.24	AL population average of Cat/Num experiment, object distrac- tor conditions at the late phase	170
J.25	AL population average of Prox/Ecc experiment, RF center at the early phase	170
J.26	AL population average of Prox/Ecc experiment, RF center at the late phase	171
J.27	AL population average of Prox/Ecc experiment, RF periphery at the early phase	171
J.28	AL population average of Prox/Ecc experiment, RF periphery at the late phase	171

J.29	AL population average of Prox/Ecc experiment, distractor only control, RF center at the early phase	172
J.30	AL population average of Prox/Ecc experiment, distractor only control, RF center at the early phase	172
J.31	AL population average of Prox/Ecc experiment, distractor only control, RF periphery at the early phase	173
J.32	AL population average of Prox/Ecc experiment, distractor only control, RF periphery at the early phase	173
J.33	MF population average without latency subtraction, Cat/Num experiment, face distractor conditions at the early phase	174
J.34	MF population average without latency subtraction, Cat/Num experiment, face distractor conditions at the late phase	175
J.35	MF population average without latency subtraction, Cat/Num experiment, object distractor conditions at the early phase . . .	175
J.36	MF population average without latency subtraction, Cat/Num experiment, object distractor conditions at the late phase	176

List of Equations

1.1 Basic formula of Divisive Normalization.	13
2.1 FOB repeat number calculation	32
2.2 RF mapping repeat number calculation.	33
2.3 Cat/Num experiment repeat number calculation.	37
2.4 Prox/Ecc experiment repeat number calculation.	37
3.1 Rational fit function for regression analysis	50
5.1 Divisive Normalization pure prediction model.	79
5.2 Divisive Normalization post-diction model.	83
5.3 Rewriting the model equation to highlight the robustness index. . . .	83
5.4 The robustness index.	85
5.5 Simple pure prediction model for for face distractor conditions. . . .	88
5.6 Simple pure prediction model for for object distractor conditions. . .	88
A.1 Definition of RF index.	133
B.1 Model with internal variables.	137
B.2 Definition of internal population activity.	137
B.3 Relating internal population activity to measured activity.	138
C.1 Relationship among angles and sides.	142
C.2 Derivation of the rotation angle ϕ	143
C.3 Derivation of the rotation angle θ	143
D.1 Converting pixels to degrees.	146
E.1 Finding the screen distance.	148
H.1 Finding the intercept.	154

H.2 Finding the projected point.	154
I.2 Relating eigenvalues and singular values.	156

Introduction

1.1 Faces as invaluable sources of social information

Faces are rich information sources. Not only do we recognize individuals by their face identity, we are also able to tell the focus of the attention by the direction of gaze. Facial expressions can also tell feeling, intention and thoughts [1, 2]. If someone is looking around impatiently, he must be anxious, whereas if he or she has a blank face without sweat and standing rock steady, we feel that person is quite confident (Fig. 1.1). We utilize these information so naturally that we often do not even think of it. Recognizing faces is crucial for social behavior and survival. Deficit in face recognition performance is known to have a link to several disorders such as schizophrenia, autism and prosopagnosia [3, 4].

According to the importance of face recognition, we primates have an excellent ability to recognize faces. We can appreciate this capability in a simple facial identity recognition task. In Fig. 1.2, four pictures taken from two in-



Figure 1.1: Faces are rich information sources. The person shown in panel (A) shows anxiety, fear and stress indicated by his wandering gaze, facial sweat and wide-opened eyelid. On the other hand, the person in (B) shows calm, blank facial expression from which one can feel his confidence and calm.

dividuals are presented. For most of human subjects, it is an easy task to recognize facial identities to group the top two pictures (A,C) and the bottom two pictures (B,D) together. Classification performance of human subjects marks stunning $>90\%$ [5] (but see [6]).

However, this is not a trivial recognition task at all. For computers, the top and bottom pictures in Fig. 1.2 (A,B and C,D) look more similar, because they have similar head orientation, lighting and facial features. In the field of machine learning and computer vision, this face recognition task had been daunting problem for decades, until recently few groups including Facebook team reported comparable machine learning performance as human, using deep learning algorithm [7]. However, their task involves only few aspects of face recognition¹, and many other human recognition skills still surpass that of

¹Taigmann et al. [7] showed that their model based on deep learning algorithm could solve a face identity recognition task involving faces with different head-orientation (i.e. head-orientation invariant recognition). However it remains to be tested whether their model can be applied to other aspects of face recognition such as luminance invariant recognition, facial expression invariant recognition [8], age invariant recognition [9] or context dependent

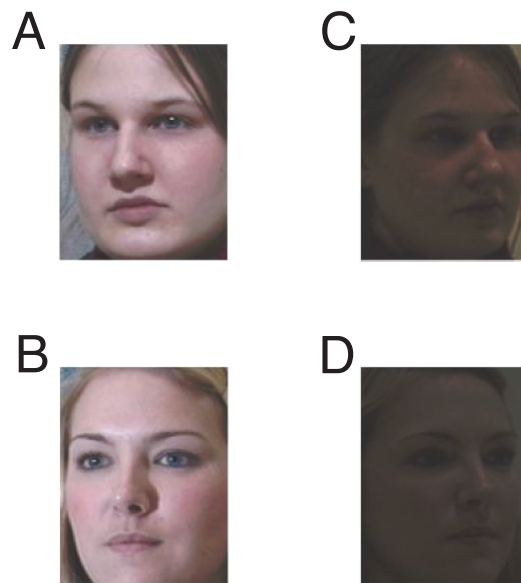


Figure 1.2: Human excels at translation invariant recognition of faces. Pictures in panels (A),(C) and (B), (D) are taken from the same individual, respectively. Even though a human subject can easily recognize individual difference to correctly group (A), (C) and (B), (D) as the same person, classification based on machine learning tended to classify (A), (B) and (C), (D) as the same group, because they have common lightening and head-orientations. Adapted from [5].

computer performance.

Similar to humans, monkeys rely on faces as information sources, and accordingly have an excellent face recognition ability. Monkeys are highly social animals and known to recognize each other's face, and communicate with facial expressions [11, 12, 13]. They utilize and recognize several facial expression patterns (e.g. fear grin, lip smacking, open mouth, cooing) to convey meanings such as fear, kinship and threat [14, 13] to other monkeys in the troop. Eye gaze also tells the focus of attention (shared attention between more than one individual by gaze following is known as joint attention [15, 16]), and failure to follow the eye can lead to impaired face recognition [17]. The fact that face recognition ability is common across primates led scientists to explore the primate brain to study where in the brain face recognition was mediated.

1.2 Hierarchical organization of the visual cortex

Sensory neurons are thought to be hierarchically organized. Each brain region such as LGN (lateral geniculate nucleus), V1 (primary visual cortex) or MT (middle temporal visual area) is thought to process different stimulus features and the selective features develop along the cortical hierarchy from low-level features such as line segment (orientation) or contrast to high-level features such as faces [18, 19, 20, 21]. In addition to the development of selectivity, a region of a field of view which can stimulate a neuron, or a receptive field (RF), is known to increase its size along the cortical hierarchy. The term receptive field was originally defined by Sherrington [22, 23, 24] and started being used for visual neurons after Hartline [25] who defined neuronal receptive field as

recognition (depending on the surrounding scene, or context, the same individual may look different to a human subject. See Sinha and Poggio [10]).

a visual field in which a stimulus can evoke activity to the neuron. Later the definition was extended to sensory neurons other than visual stimulus. However, even stimulus outside Hartline’s definition of RF (classical RF) was found to influence a single neuron response. Stimuli in these non-classical RF can modulate neural response [26, 27, 28, 29, 30], depending on stimulus context (e.g. [31, 32, 33, 34]). Models based on lateral interaction or feedback inhibition were proposed to explain the mechanism of non-classical RF [35, 36, 37], and recently it was shown in experiments that specific type of interneuron provides lateral inhibition to mediate surround suppression [38, 39]. In this thesis, we use the term RF according to Hartline’s original definition: visual field in which a single stimulus can drive an activity to the currently recorded face selective neuron.

Hubel and Wiesel proposed a model of V1 neurons to explain the mechanism of development of neural response property along the hierarchy [40]. For example, they suggested that the elongated excitatory oval shape (flanked by inhibitory region(s)) of V1 neuron’s RF and its orientation selectivity was developed by converging afferents from LGN neurons in the thalamus, whose RF have concentric, antagonistic center-surround structure (either excitatory center (on-center) or inhibitory center (off-center)). Namely, they thought superimposing LGN RFs in one particular direction could create elongated excitatory RF of V1 to have selectivity to a stimulus with particular orientation (“simple cell”). Similarly, “complex cells” in V1 which do not have clear excitatory and inhibitory regions in RF but still selective to orientation, were thought to be developed via converging afferents of V1 simple cells. Based on Hubel and Wiesel’s model of hierarchical development of neural response [40, 41, 42], Riesenhuber and Poggio developed a hierarchical model of visual object recog-

dition, where stimulus selectivity and invariance² [43, 44, 45] increased along the hierarchy. The hierarchical model contains interleaving simple cell layers and complex cell layers, which take weighted sum and max³ of incoming afferents, respectively. As a result of repeating weighted summation and max operation, both selectivity and (translation and scale) tolerance increase along the hierarchy.

1.3 Neurophysiological evidence of face selective neurons in the temporal lobe

The first evidence of the neural representation of face was provided by Charles Gross and his colleagues in 1972 [46, 47]. They conducted electrophysiological recordings in the macaque brain, in the superior temporal sulcus (STS) where object selective neurons reside in [48, 19, 49, 50, 51, 52, 53]. They found that neurons in the temporal lobe responded preferentially to a facial image, compared to other non-face images. After the discovery, face neurons were also found in several other regions in STS [54, 55, 56] and also in different cortical and subcortical regions including amygdala [57, 58, 59, 60], hippocampus [61], pulvinar [62], entorhinal cortex [63, 64] and orbitofrontal cortex [65].

In those days - from the 60s to the early 90s - finding a face selective neuron⁴ was difficult and less reliable experiment. Researchers started knowing that temporal lobe contained visual neurons selective to complex images [19, 20, 21],

²Neural response is called as invariant if the response is unchanged to translation of stimulus, such as stimulus scaling, rotation (2D or 3D) or contrast change. The term invariant is also applied to visual recognition if the recognition is not altered by translation of stimulus (e.g. head-rotation invariant individual recognition).

³Thus, a “max neuron” responds as if it ignores all presented stimuli except its preferred stimulus.

⁴In this thesis, unless otherwise noted the term “face selective neuron” is used to mean either implicit indication or explicit quantification of the following: the neuron respond at least twice as strongly to the face as to other non-face objects [67, 68, 69].

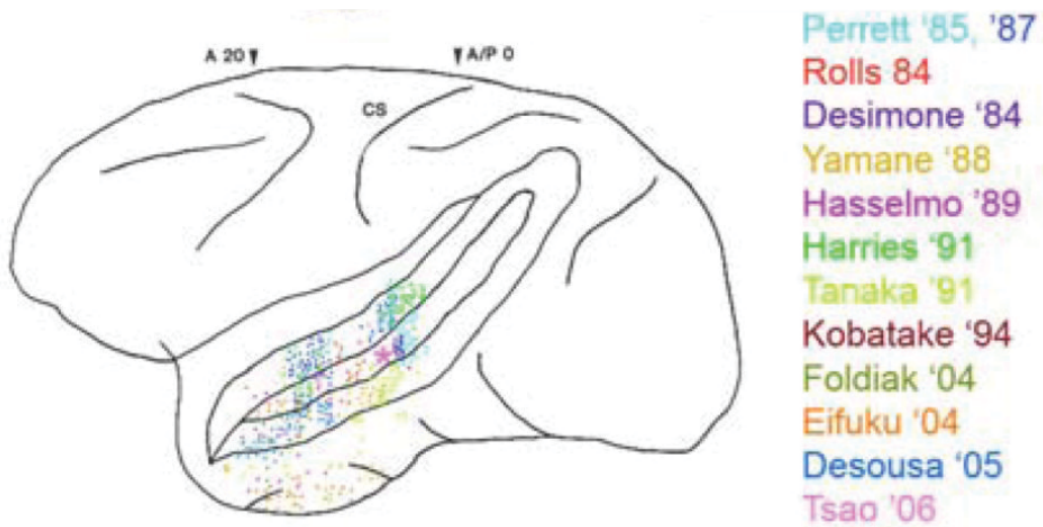


Figure 1.3: Face selective neurons found in the macaque temporal lobe. Summary of discovered face selective neurons in 80's - 00's is shown. Without having a prior knowledge of the location of face neuron clusters, finding face selective neurons were risky, difficult experiment. Recently, Tsao and Livingstone [66] compared the distributions of reported face selective neurons (shown above) and functional imaging data to find that the physiological data had high concentration in two regions, potentially corresponding to middle and anterior face patches (see the text for details) reported by Tsao et al. [67]. Adapted from [66].

but whether these object selective neurons formed cluster or not was unclear, and had been a controversial topic [70, 71, 72]. Thus, to find face selective neurons they had to penetrate electrodes according to the previous literatures and had to test every neuron they encountered whether it was a face selective or not. This procedure took excessive amount of time. Different researchers reported face selective neuron in various part of temporal lobe (Fig. 1.3, [73, 74, 75, 76, 57, 77, 78, 79]), making it difficult to decide which brain region should be recorded.

However in the 90s, thanks to the advent of functional magnetic resonance imaging (fMRI), researchers were allowed to image the entire brain to localize face selective cortices. Although fMRI measures brain activity by detecting blood flow⁵ which has low spatiotemporal resolution, it helped researchers to localize a functionally distinct brain region activated by particular stimulus or cognitive state [80, 81]. Aided with fMRI, Nancy Kanwisher and colleagues found what they called fusiform face area (FFA) in the human temporal lobe⁶ [83, 84, 85]. When a human subject viewed an image of a face, the FFA increased its activity, measured by the blood-oxygen-level (BOLD) contrast (Fig. 1.4). Later they also found that brain regions selectively engaged when a human subject perceives body, place or words were also segregated [70, 86]. These results supported the view that visual objects of different category are processed by functionally distinct modular structure. Finally Doris Tsao, Winrich Freiwald and colleagues found six and three segregated face selective cortices, or “face patches” in the macaque temporal and frontal lobe⁷, respectively [67].

⁵Blood provides energy and nutrition to highly activated brain region, which can be used as a surrogate of brain activity.

⁶But see also [82] for the earlier positron emission tomography (PET) study indicating the existence of the functional organization of face and object processing cortex.

⁷They also compared macaque and human brain using fMRI to find homologous face patches along the temporal lobe [87].

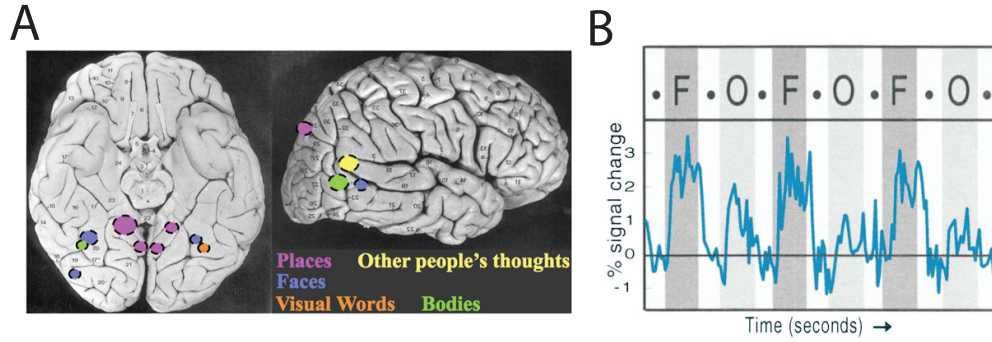


Figure 1.4: Face selective region in human temporal lobe revealed by fMRI. (A) The regions activated when a subject perceives a certain category of stimulus are highlighted. Fusiform face areas are depicted the purple circles, located bilaterally in the midfusiform gyrus. The activated regions were determined by the level of BOLD signal indicating high energy consumption. Adapted from [70]. (B) Fusiform face area was selectively activated by facial image (F) but less by non-face image (O), in the human temporal lobe. Adapted from [83].

1.4 Face selective neurons are segregated into modular structures

Of the six temporal lobe face patches, the most posterior one is located at the area TEO (posterior part of inferior temporal (IT) cortex), named PL (posterior lateral). The other five patches are located at the area TE (anterior part of IT). The two posterior “middle face patches” were named ML (middle lateral) and MF (middle fundus), after the location of the STS. Two other patches, also located in lip and fundus were named AL (anterior lateral) and AF (anterior fundus), respectively. The anteriormost face patch located in anterior middle temporal sulcus (AMTS) was named AM (anterior medial) [88] (see Fig. 2.7). Face patches were not an exception of the hierarchical organization of sensory cortices: with fMRI-guided microstimulation and histology, these face patches were shown to be interconnected reciprocally [88], and each face patch contained face selective neurons with different response characteristics. Thus, the

macaque brain has a dedicated neural circuit for processing and transferring information⁸ of a face along a cortical hierarchy. For example, middle face patch (MF, ML) neurons are selective to head-orientation, a large fraction of anterior patch (AF, AL) neurons respond to either left or right profile view (i.e. mirror-symmetric tuning), and anterior-most patch (AM) neurons respond to particular facial identities [89] (a recent study also indicated that the posterior face patch (PL) neurons were selective to eye region [90]). Thus, face selective neurons were found to create a mutually connected modular structure thought to mediate primate face recognition ability.

The fMRI-guided electrophysiology facilitated studying cortical information processing and circuit structure: we acquired precision to target our region of interest, with a prior knowledge that more than 90 % of the neurons in the region are face selective [67]. This reduced our need to explore neural selectivity extensively, allowing us to start testing experimental hypothesis right after we reached to the target face patch⁹.

1.5 Early evidence of multiple stimulus representations and proposed models

In a typical, classic vision study researchers used only one visual stimulus on a gray screen, on which a subject monkey needs to fixate [49, 19, 91, 92, 93]. In the real world, however, we encounter numerous visual stimuli at a time: an image of face might be surrounded by other faces or non-face

⁸In this thesis, I use the term “information” as a quantity conveyed by neural activity, from which an observer can infer stimulus of outside world or decode an internal variable such as memory or attention. Thus, it does not necessary mean Shannon information or Fisher information *per se*.

⁹Without prior knowledge of neural selectivity, a researcher of face neurons needed to verify that the neuron being recorded was face selective. Although we still conduct an experiment to test neural selectivity, we could narrow down our exploration space to focus on our main interest.

objects. Still, however, primates are capable of recognizing faces even in highly cluttered natural scenes [94]. Thus, in order to understand how the primate brain processes faces in natural scenes, using multiple stimuli to characterize response property of face selective neurons is crucial. Throughout this thesis, I use the term "stimulus representation" to refer neural activities from which an observer can infer the presented stimuli. Neural activities might be a single neuron response, or representation space spanned by neural population.

Several studies gave hints to the question of how visual neurons respond to multiple visual inputs, by presenting two stimuli in neuronal RFs of either the visual area V4 or MT [95, 96, 97, 98]. Reynolds et al. found that without visual attention the response to neuron's preferred stimulus and non-preferred stimulus (both inside the RF) was approximately an average of the two stimuli. Moreover, the response could be biased by instructing a subject to attend to one of the stimulus, leading to larger response (attended to preferred stimulus) or smaller response (attended to non-preferred stimulus). Based on this observation, Desimone and Duncan proposed the Biased Competition model, which states that top-down (and/or bottom-up) attention can bias the competition between two stimulus representations over neural resource [99]. These visual attention studies indicated that neural response to multiple stimuli were an average of the responses to isolated stimulus presentations if no visual attention was exerted, which tempted researchers to study whether this regime of integrating multiple inputs was universal across visual cortex. Additionally, since temporal lobe neurons have much larger RFs which can encompass more than two stimuli at a time, it was important to know if the average operation applies to three, four or even more stimuli.

In order to explain cortical normalization effect including averaging operation, David Heeger and colleagues [100, 101, 102, 103], developed the Divisive

normalization model that is now widely used in the neuroscience field, especially sensory integration researches [104, 105, 106, 107, 108, 109, 110, 111, 109, 112, 113]. Divisive normalization states that a neuron computes a ratio between the direct inputs and population activity. The concept is illustrated in Fig. 5.1: a neuron receives inputs from visual stimuli, and neighboring population of neurons are also receiving inputs from the same visual stimuli. Then, through lateral interactions or converging afferents, the neuron’s activity will be normalized in proportion to the population activity of neighboring neurons. Response normalization by lateral interaction *in vivo* was recently shown experimentally by Matteo Carandini in mice V1 [114], and by John Reynolds in macaque brain [115], both using optogenetics.

Divisive Normalization can be integrated into the Biased competition model proposed by Desimone and Duncan ([99], see Chapter 1). Biased competition states that attention induces top-down modulation from cortices in upper hierarchy (e.g. frontal eye field [116]), which biases the allocation of neural resource to competing multiple inputs. Similarly, bottom-up attention (e.g. pop-out effect, see Chapter 1) can bias the allocation of resource. Divisive Normalization can also bias which input dominates the final output by population activity of neighboring neurons¹⁰. Or, alternatively, all inputs could have equal privilege, resulting in averaging all the inputs. In fact, Divisive Normalization can be applied to Biased competitive model with a common mathematical formula¹¹, as reviewed by Carandini and Heeger [103], and Reynolds and Heeger [111]. The common formula is shown in Equation 1.1:

¹⁰For example, high contrast stimulus takes more neural resource than low contrast stimulus. Similar to this idea of Biased completion model, Divisive Normalization states that high contrast elicits large activity to the neighboring population to have larger weight (i.e. bias) to the high contrast stimulus. See [112, 103].

¹¹Note that Equation 1.1 is a general form, and details can be different in some studies, such as free parameter in different forms [109, 112] or omitting negligibly small terms [103, 111].

$$R^{1,2} = \frac{c1 \cdot I1 + c2 \cdot I2}{c2 + c1 + \sigma} \quad (1.1)$$

where $R^{1,2}$ is a neural response to stimuli 1 and 2, $c1$ and $c2$ are coefficients, $I1$ and $I2$ are inputs (thus, Equation 1.1 is modeling a response to two stimuli. It can be easily expanded to more than two stimuli by having as many terms as stimulus number) σ is a spontaneous activity of a neuron¹². The two coefficients $c1$ and $c2$ can be functions of population activity (in Divisive Normalization), of top-down attention (in Biased competition) or both (see [109]).

1.6 Evidence of multiple stimulus representations in IT cortex

Zoccolan et al. (2005) [107] conducted recordings with up to 3 stimuli simultaneously presented. They recorded from central to anterior temporal cortices while subject monkeys were passively fixating on the center of a stimulus presentation screen. They concluded that the neural response to multiple stimuli were predicted well by the average of the responses to individually presented stimuli, regardless of RF location (did not depend on effective position inside RF), and regardless of stimulus identity (i.e. the response depended on the response magnitude, not on stimulus identity). However, they relaxed and extended the argument two years later. Zoccolan et al. (2007) [117] trained subject monkeys for an object detection task and presented neuron’s preferred stimulus with various transformations such as position, size, contrast changes

¹²In the following modeling works we did not use the σ term because we subtracted spontaneous activity from the data used in simulations.

and clutter addition (they used up to two stimuli this time: one clutter out of pre-selected 6 non-preferred stimuli was used). During the object detection task, the subject was required to maintain fixation on a center dot, until they found the target (a red triangle) to make a saccade to a fixed location to receive rewards. They reported a trade-off between translation invariance and stimulus selectivity: namely, if a neuron had a sharp tuning curve selective to only one or few stimuli, the neuron was sensitive to stimulus translation or addition of clutter (i.e. decreased the mean firing rate). However, if a neuron had a broad tuning curve, the neuron was robust against stimulus translation and addition of clutter didn't reduce firing rate either. Thus, neurons with broad tuning curves responded to multiple stimuli with higher magnitude than average operation predicted. They ascribed this apparent discrepancy to difference in selection criteria of neurons: in Zoccolan et al. (2007), more inclusive criteria was used, resulting in collecting neurons with broad tuning curves too. They explained the tradeoff between stimulus selectivity and invariance by presenting two descriptive, qualitative models. In a model what they called "toy model", neurons were assumed to have multidimensional Gaussian tuning curves in a input stimulus space. Sparse responding neurons had Gaussian tuning curves with smaller standard deviation, thus the (multidimensional) area providing an effective input is smaller than broad responding neurons. Because of small effective input space, the sparse responding neurons had less tolerance to the transformations. This model captured a qualitative reduction of tolerance as stimulus selectivity increased. In their second model, they applied the hierarchical model of object recognition, originally proposed by Riesenhuber and Poggio (see above, [43, 44, 45]). The hierarchical model reproduced Zoccolan et al's result qualitatively: however, quantitative prediction of a response of a particular neuron remained to be studied. Also, they did

not necessary record from a cortex with specific object selectivity, which could potentially affect their results¹³.

Rolls et al. [118] recorded from IT cortex neurons using one or two stimuli on a complex natural background. Since the natural background could contain multiple objects (e.g. leaves, trees), within the RF of IT neurons multiple objects would be encompassed. During the recordings, they imposed a visual search task: the subject needs to search for one of the two stimuli presented to touch to receive rewards. As a result they found the response of a neuron was enhanced when a preferred stimulus was inside the RF and the stimulus was a search target. In addition, they reported that when the effective stimuli were presented on the complex natural background, RFs of the IT neurons became smaller than presented on a plain background, regardless of attention (i.e. regardless of whether the preferred stimulus inside RF was a search target or not). They claimed that by reducing the size of RF, IT neurons could reduce confounds introduced by multiple objects inside the receptive field. The RF shrinkage could be due to an iceberg effect due to overall reduction of the firing rate across visual field with cutoff at firing threshold, but detailed mechanism remained unclear. Rolls et al. defined RF as the visual field elicited firing rates above the baseline when it contained the preferred stimulus. The RF was measured in distance (visual degrees) between the eye position and the stimulus, (not degrees from the fovea), and they reported almost monotonic decrease of activity as the stimulus was drawn away from fovea (i.e. when the stimulus occupy the center vision, the neuron was always activated), regardless of background and attentional condition. This either implicitly assumed all the

¹³At first sight, our experimental results looked contradictory to Zoccolan et al.'s studies (see Chapter 3). However, we built a quantitative and mechanistic model to explain the different results by difference in cortical structure: because we recorded from face selective cortices, and because we compared face and non-face stimuli, the result was seemingly different. However, our computer model could give a unified framework explaining both our result and Zoccolan et al.'s result. See Chapter 7.

recorded neurons have foveal RF, or they happened to record from a cortical region, all of whose neurons have foveal RFs. In our recording, most of the RFs had positional bias toward a visual field contralateral to the recording hemisphere. Thus, direct comparison of our results and that of Rolls' et al. might be difficult. There are few more differences between our experiment and Rolls et al.'s. Firstly, in all conditions Rolls et al. used visual search tasks that involve attention. Secondly, their stimulus presentation time varied: they selected time period when the eyes were still (within 1 degrees of visual angle) for ≥ 100 ms during a free view, and calculated firing rate for each 100 ms period. Thirdly, Rolls et al. (and Zoccolan et al. too) considered stimulus (or stimuli) inside the RFs, and it was not clear how a neural response was affected by stimuli outside of RF (Does a neuron completely ignores the stimuli outside of its RF, or do the stimuli still affect neural responses, probably through non-classical RF?). Thus, although we know some of properties by which a neuron integrates and represents multiple stimuli, small differences across experiments make it difficult to have direct comparison or to have a unified understanding of computational principle of temporal lobe neurons.

1.7 Multiple stimuli representations in the macaque temporal lobe face patch

As I reviewed above, we have limited knowledge about visually selective neurons in primate brain. Then, how do neurons with a specific selectivity - face selective neurons - integrate and represent multiple inputs from visual stimuli? Do they compute an average of all inputs, take max of them, summate them or could it be nonlinear summation? If the neuronal firing rates are an average of responses to individually presented stimuli, is a representation of a preferred

face degraded with presentation of non-preferred stimuli because of a response reduction? Or, alternatively, if neurons respond to the most preferred face while ignoring all other non-preferred stimuli, i.e. do not reduce response with non-preferred stimuli, do they discard all the information about the presented non-preferred stimuli? Is the neural response dependent on distractor parameters such as distractor number, distractor category or distractor proximity? And, if so, can an observer read out information of outside world (i.e. what is presented on the screen) from single neuronal firing rates and/or from population activity patterns? In the following chapters, I will address these questions with experimental results and a computer model to suggest potential circuit mechanism underlying multiple stimulus representation.

1.8 Organization of the thesis

This thesis is organized as follows. Chapter 2 explains the details of the experiments including targeting the face patch, electrophysiological recording technique, stimulus design and flow of the experiment. In Chapter 3, I describe the basic response property of the middle face patch, MF, mostly by analyzing mean firing rates of single cell and population. Then I raise the question to analyzing data just based on the mean firing rate, leading into Chapter 4 where I investigate stimulus representation in an entire population of neurons, using machine learning, dimension reduction technique and the independent component analysis. I explain how the heterogeneous yet systematic response of the neurons carry detailed information that is not seen just from mean firing rates. In Chapter 5 I introduce a computer model based on Divisive Normalization framework to address how a modular structure and functional connectivity of the face selective temporal lobe neurons can confer discriminability to neurons. I also describe the way neuronal responses become

more robust against distractor stimuli over time, using robustness indices derived from the model. Chapter 6 compares the MF data with that of anterior face patch (AL), which is located at one step up in the hierarchy of the face patch network and has slightly different response property from MF neurons. Finally in Chapter 7, I summarize the findings of the project and highlight the importance of modularity, heterogeneity and response systematicity for representing multiple stimuli in the population of face selective neurons in the macaque temporal lobe. I also compare the current study with previous literatures to argue that our computer model can give an unified account to previous reports which are seemingly contradict each other.

2

Experimental design

2.1 Varying distractor numbers to study neural computation of multiple visual inputs

In order to test either a face selective neuron responds as an average of multiple inputs or as a max of them (or some other responses such as a summation or a nonlinear function of the stimuli), we had to select and combine a stimulus elicits a large response from a neuron, and a stimulus that elicits a minimal response from the neuron. This is because, for example, if two images elicit exactly same magnitude of responses to the neuron, one cannot discriminate whether neurons are taking an average or max of them (imagine an input vector $(1,1)$ - the average of it is 1, and max of it is also 1). Natural scenes containing tens or hundreds of objects were also not suitable stimuli to investigate computation, because (a) we did not know how strong a drive each object gives to a single neuron, and (b) we did not know how objects interact together to affect a neural response. Therefore, we decided to use a preferred

face, a non-preferred face (distractor¹) and combination of them to study neural representation of multiple stimuli. If a neuron is taking an average of all the visual inputs, adding non-preferred stimuli would result in reducing firing rates. On the other hand, if a neuron takes maximum input while ignoring all other ineffective inputs, the firing rate may not be reduced by increasing number of distractors. Thus, the first stimulus parameter we decided to use was the “distractor number” (Fig. 2.1). We placed the preferred face at the RF center, while presenting up to eight non-preferred face in the surrounding position² in order to study whether the computational rule is consistent across different number of stimuli.

2.2 Varying distractor categories in light of feature / conjunction search

Adding more than two distractors was a natural extension of Zoccolan et al.’s research to study how a face selective neurons compute afferents driven by visual stimulus. This “computation” - average, max, summation or nonlinear response to multiple inputs - could depend on the context of surrounding distractors, similar to surround suppression or contextual modulation by non-classical RF [34, 36, 119, 120]. Additionally, in light of pop-out effect in visual search, whether the preferred face and distractor have common feature or not may influence neural response. Pop-out effect is used to describe the fact that the reaction time and number of items on the search screen (set size) are independent (set size v.s. reaction time plot becomes flat), and the fact that

¹In this thesis, I use the term a “distractor” same as a “non-preferred stimulus for a given neuron”.

²We also had a control stimulus condition which had the preferred face at one of the surrounding positions while the center position is occupied by a distractor. for details, see Chapter 2.



Figure 2.1: Distractor number variation. Distractor stimuli, which by themselves elicit minimal activation to the neuron are placed at up to 8 positions around the neuron's most preferred stimulus. Distractor stimuli were placed in point-symmetric position. See text for details.

the compelling phenomenological impression that the target is immediately visible [121, 122]. Treisman and Gelade described this effect in their Feature integration theory [123, 124]. According to the Feature integration theory, searching for a target among distractors which do not have common feature is conducted with a fast, parallel process (thus, the target “pop-out” among distractors). The feature can be low-level features such as line segment or color, or alternatively a high-level feature such as a face [122]. For example, an upright face can pop-out among inverted faces even if all the faces share similar colors [125]. Treisman and Souther proposed that the requirement for being distinct features is to be represented in non-overlapping neuronal groups [126]. This parallel process was defined as “Feature search” and it was shown that reaction time to find a target among distractors was almost independent of distractor numbers: no matter how the number of distractors was increased, the time required for finding a target was almost constant. In contrast, if distractors have similar features in common (e.g. color or shape), the search becomes serial search, whose reaction time becomes proportional to the distractor numbers. Thus it was defined as “Conjunction search”.

Fig. 2.2 demonstrate feature search. In this figure, a person - Peter Higgs - is standing in front of the Large Hadron Collider, or LHC, which verified the existence of Higgs boson originally proposed by him. This picture is quite busy, or crowded, in sense that Dr. Higgs is surrounded by miscellaneous mechanical parts like cables, metal plates and beams. Nonetheless, we can perceive Dr. Higgs’ face almost immediately and automatically - this is the pop-out effect. According to the Feature integration theory, Dr. Higgs’ face pops-out from the picture because surrounding objects do not have common feature with the face³. In contrast, Fig. 2.3 demonstrates Conjunction search. In this case an

³Hereafter I call the objects which do not have common feature as objects in “different category”. For example, face, fruit, letter are different object categories.



Figure 2.2: Example of Feature search. Peter Higgs is standing in front of the Large Hadron Collider. Despite the intricate surrounding objects, we can easily recognize his face in the picture.

observer experiences a totally different perception - to recognize one particular face out of the crowd, an observer needs a long time to scan faces one by one until he or she finally finds the face in the crowd.



Figure 2.3: Example of Conjunction search. Perception is depending on the surrounding context: in this picture, the same face of Peter Higgs as in Fig. 2.2 is embedded but it is extremely difficult to recognize his face.

What is the neurophysiological basis of feature / conjunction search? During feature search, where the target stimulus pops out independent on the number of objects in the display, a subject can find a target in almost constant reaction time. Thus, one prediction is as follows: there are neurons which do not reduce their mean firing rate⁴ by presenting distractor objects in addition to their preferred face. Such neurons look as if they take maximum input (from the preferred face) while ignore all other small inputs provided by distractor object⁵. On the other hand, in conjunction search, where the target stimulus share feature with distractors, the mean firing rate of neurons might

⁴For simplicity, mean firing rate of single neurons are mentioned here but the stimulus information can also be coded within an entire population (population code), which we actually found. In this case, a stimulus could be decoded from the response pattern of the neural population, even though mean firing rate of each neuron can change.

⁵Such “max neurons” are also requirements of the aforementioned hierarchical model of object recognition proposed by Riesenhuber and Poggio.



Figure 2.4: Distractor category variation. Either face or object distractors were presented in surrounding positions. Note that distractors themselves (i.e. without the preferred face) elicit minimal response from a neuron.

strongly depend on the number of distractors on the screen.

The neurophysiological correlates of pop-out effect starts from striate cortex or V1 [127, 128]. The response to an oriented bar was reduced when flanking oriented bars were presented together, and the neural response was larger when flanking bars had different orientation from the center bar (pop-out condition) than when the flanking bars' orientations were the same as the center bar. This neurophysiological pop-out effect seems to continue in higher visual areas too. Beck and Kastner conducted human fMRI study involving a searching task among multiple stimuli where a target pops out (distractors do not have common feature as the target), or does not pop out (distractors have common feature to the target [129], see also [130, 131]). They found BOLD response to a non-pop out stimulus showed a decrease, indicating competitive interaction among stimuli (also in the framework of Desimone and Duncan's biased competition model), whereas the pop-out stimulus did not show competition. Even though our experiment did not involve a searching task, neural activity might be affected by common feature among distractors and preferred

stimulus. In light of this perceptual difference affected by distractor context, we decided to use distractors with different features, namely face distractors or object distractors (Fig. 2.4). We selected face and object distractors which evoked almost no response to the cell and used one of them to study the effect of distractor category to neural representations.

2.3 Varying distractor proximity / eccentricity in light of the crowding effect

The arguments in the previous section are closely related to the “crowding effect” found in human psychophysics studies [132, 133, 134]. When a human subject is instructed to identify a target stimulus with flanking distractors, the subject can detect the target yet finds it difficult to identify (Fig. 2.5) when the distractors and targets share common features. This feature does not have to be a low-level feature such as a line segment, color or Gabor but can also be a high-level feature such as object category. For example, face distractors can induce crowding effect, but inverted face distractors reduce the effect [135]. The crowding effect is absent at the fovea but becomes stronger at larger eccentricity (i.e. more peripheral vision) and at closer distractor proximity, a property known as Bouma’s law [136]: the minimal distance allowing subjects to identify the target scales with stimulus eccentricity, although a few exceptions are reported [137, 138, 135]. There is an apparent relationship between the crowding effect and the aforementioned Feature integration theory: if distractors and a target have a common feature, the crowding effect is induced and an observer needs to move the eyes to place one object to the fovea at a time. That’s why the reaction time to find the target is proportional to display size, as the conjunction search in Feature integration theory explains.

Alternatively, if there is no shared feature between distractors and a target, the crowding effect is not induced and an observer can identify the target even in the peripheral vision. This quick, parallel search is stated as a feature search in Feature integration theory and demonstrates that the reaction time to find the target is almost constant (i.e. pop-out) across large variation of display size.

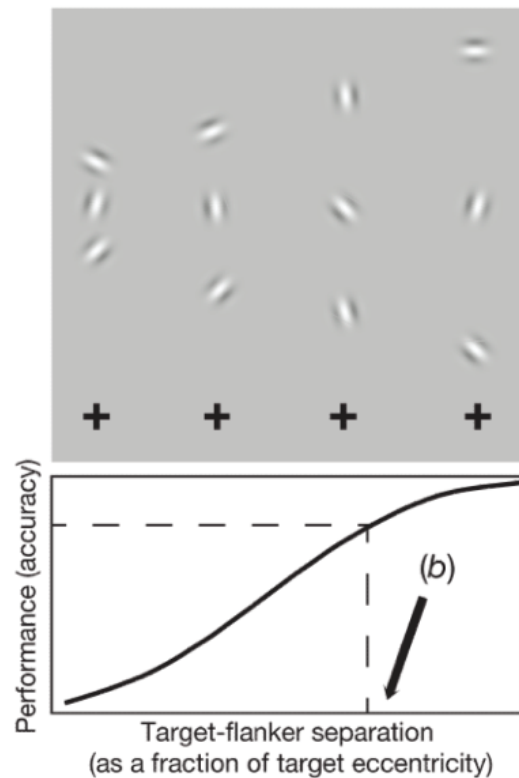


Figure 2.5: Psychophysical crowding effect. Human subjects were asked to fixate on the black cross, while being asked to identify direction of the center flanker. As the top and bottom flanking bars move close to the center bar, the identification performance drops. This effect is known to be even stronger at the peripheral vision. Bauma’s law states that the distractor distance which allow a human subject identify the center target, or “critical distance”, is approximately equal to the stimulus eccentricity. Adapted from [135]

Several mechanisms subserving the crowding effect were proposed. One of them is a pooling model, in which local features are averaged to reach perception rather than a target feature being masked by distractors [139, 140, 141, 142, 143]. An observation that supports this model is that human subjects are capable of reporting average Gabor angle, even under the crowding condition. Physiological observations also support the pooling model. RF size increases throughout the visual cortical hierarchy, which suggests large integration over visual field resulting in some information loss [144]. Additionally, the existence of lateral interactions and mutual inhibition also indicates local interaction among neurons with overlapping RFs [145, 146, 147]. Thus, the pooling model states that the local features were pooled or averaged within a certain visual field in the peripheral vision. Freeman and Simoncelli demonstrated this hypothesis by presenting “metamers”, which are different images sharing common statistics [148]. They showed that when the pooling region size of the model was set correctly, human subjects could not distinguish two metamers in peripheral vision, although subjects could appreciate the difference of metamers when they fixate onto the metamers to bring the images to center vision. Rosenholtz et al. also created image patches (they call them “mongrels”) with same summary statistics (e.g. correlation of responses of V1-like orientation tuned neurons across locations / luminance autocorrelation / marginal statistics of luminance / phase correlation, etc.) to show the human subjects could not distinguish the image patches with losing local information such as position of line segment [149]. Substitution was also suggested as a mechanism of crowding model [150, 151, 152]. The model states that a target object is often substituted or confused with one of flanking objects, resulting in failure in reporting correct identity of a target object. Another model proposed by Dayan and Solomon utilizes Bayesian inference [153]. Bayesian inference is

a method to give an estimation based on minimization of a loss function which expresses the cost of estimation errors [154]. In their model, the crowding effect occurs by a target selection among flankers emerging through Bayesian inference within large RF with spatial uncertainty. The spatial uncertainty induces interference between stimuli in neighboring space that affect inference. Even with ample of suggested mechanisms, however, none of the existing models can explain the crowding effect perfectly and it remains a matter of debate [135]. Some of models require fine-tuning of parameter to reproduce psychophysics result, or others explain a part of the crowding effect without clear explanation to other parts such as tangential anisotropy (radially positioned flankers are more effective than tangentially positioned ones [137]).

In light of the psychophysical crowding effect, we introduced the third and forth stimulus parameters: distractor distance and distractor proximity. In this stimulus distractor numbers were fixed to eight (to reduce the number of combinations), while distractor distances were changed from 0, 2 to 4 degrees of visual angle (Fig. 2.6). These stimuli were presented either at the center of the measured RF, or 5 degrees away from the RF center toward peripheral visual field. For details of stimuli, see Chapter 2.

2.4 Stimulus selection and composition of multiple stimuli

All the experimental results presented in this thesis were recorded from the middle face patches MF (medial fundus, Fig. 2.7) and anterior patches AL (anterior lateral). Neurons in the face patches are face selective ($>90\%$, see [67]), and MF/AL neurons are tuned to a head-orientation of face image [89]. This known response property allowed us to select highly effective preferred face

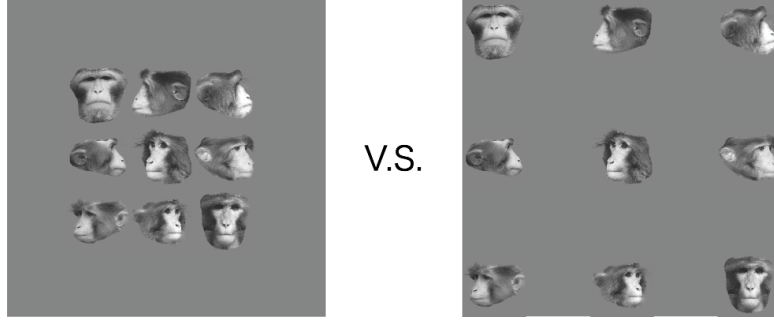


Figure 2.6: Distractor proximity variation. distractor proximity was changed such that vertical and horizontal separation between the preferred face and the distractors were from 0, 2 (not shown) to 4 degrees. We also changed the stimulus position, either the center of the measured RF or 5 degrees away from RF center (in the peripheral visual field)

(face with preferred head orientation) and ineffective, non-preferred face (face with non-preferred head orientation). AL neurons are also tuned to head-orientation, but many of them are tuned in mirror-symmetric (i.e. respond both to left and right side view) manner. In both patches, very few of them are tuned to particular individuals, unlike in AM (anterior medial). Instead, MF/AL neurons are in general tuned to one or two head-orientations and respond to several different face identities. Few of neurons responded to a particular individual among our stimulus set. These highly selective neurons are often sparse responding with low spontaneous firing rate.

On each neuron we encountered, multiple experiments were run. The initial experiment was to find a preferred face, non-preferred face (face distractor) and non-preferred, non-face object (object distractor), in order to have sufficient dynamic range to test either a face selective neuron takes average, max or summation of inputs. To select the three images, we presented 147 different

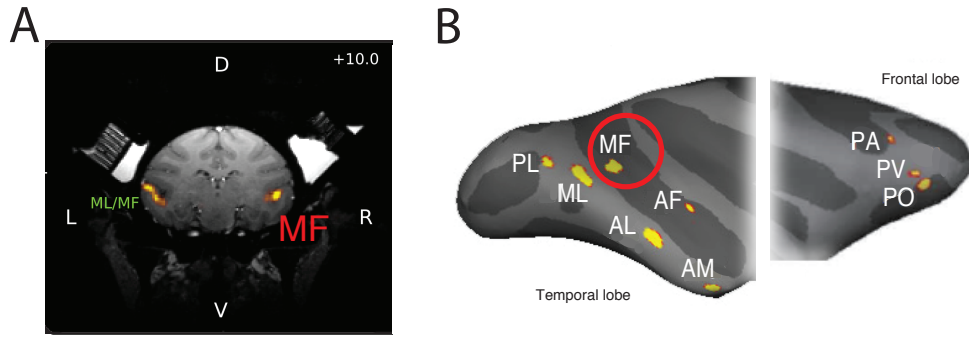


Figure 2.7: The target face selective cortex, MF. The panel (A) is a coronal section showing the middle face patch, MF. Shown is functional MRI image overlaid with structural MRI (T1 weighted), taken from one of the two monkeys recorded (monkey J). The middle face patch MF locates bilaterally in the fundus of the superior temporal sulcus (STS). The panel (B) is an inflated map of macaque brain, showing 6 temporal and 3 frontal patches. Both temporal and frontal patches are bilateral and known to have reciprocal connections to form face processing network [88]. The vertical separator indicates that the data of temporal and frontal patches are taken from two distinct studies [67, 155].

images containing face, object and body (FOB stimulus) while recording from a single unit. 80 of stimuli were face images, consisted of 40 human faces and 40 monkey faces (because most neurons were selective to either human or monkey faces, we included both macaque faces and human faces). For each species, 8 individuals with 5 head-orientations were used: $-90/90$ degrees (full profile view), $-45/45$ degrees (half profile view) and front view. 66 non-face object stimuli consisted of technological objects, place views, fruits, monkey bodies, human bodies and human hands, each of them containing 11 different identities. Gray square with identical color as background (R128/G128/B128) was also included as a control stimulus. Each stimulus was presented for 200 ms per image without inter-stimulus interval, and the image size was fixed at 4 degrees of visual angle. The image order was randomized and presented for 2.5 minutes to record approximately 5 repeats per image (calculation shown in Equation 2.1).

$$\frac{200[ms/image] \cdot 147 \cdot x[image]}{10^3[ms/sec] \cdot 60[sec/min]} = 2.5[min] \quad (2.1)$$

$$x \approx 5.10[repeats]$$

Stimuli were presented at a putative RF center, which was determined by moving stimulus position manually. Subject monkey was required to keep fixation on a center dot of the presentation screen 57 cm away from the eyes⁶ for 3-5 seconds in order to receive liquid reward (either water or juice). Eye position was measured by using infrared pupil tracking system (ETL-200, ISCAN Inc., Burlington, MA), and as soon as the eye position deflected more than 3~5 degrees away from the fixation dot, reward stopped immediately. Mean firing

⁶57 cm separation between eyes and a presentation screen equalizes 1 cm on the screen to 1 degree of visual angle. For details, see E.

rates to each stimulus were calculated online using custom MATLAB script, and the preferred face, non-preferred face and non-preferred object were chosen based on the firing rate (Fig. 2.8A) in order to use for following experiments.

Next, using the selected three stimuli we mapped the neuronal RFs. One of the three selected stimuli were presented at one location of hexagonal grid pattern (the grid was invisible, in Fig. 2.8B the grid is shown for explanation purpose) on the stimulus presentation screen for 200 ms without inter stimulus interval. The grid was spanning 18 (vertical) \times 24 (horizontal) degrees of visual angle (presentation locations separated by 3 degrees), covering almost the entire screen (68 presentation positions in total). RF mapping stimulus was presented for 6.5 minutes, resulting approximately 10 repeats per stimulus presentation position, as shown in Equation 2.2. Fig. 2.8B shows the three mapped RFs of an example neuron. All of the subsequent stimuli were presented at the center of the RF of preferred face, unless otherwise noted.

$$\frac{200[ms/image] \cdot 3x \cdot 68[image]}{10^3[ms/sec] \cdot 60[sec/min]} = 6.5[min] \quad (2.2)$$

$$x \approx 9.56[repeats]$$

The remaining experiments were conducted to test multi-stimulus integration with variables introduced in Chapter 1: distractor number, category and proximity. Fig. 2.9 shows the detail of the multiple stimuli we used for distractor category / numbers variation experiment (Cat/Num) and the distractor variation experiment (Prox/Ecc). In both experiments, the preferred face was placed at the center (except the control stimuli, see below). For the distractor numbers / category variation experiment, 1, 2, 4 or 8 face distractors or 1, 2, 4 or 8 object distractors were placed around the preferred face. distractors

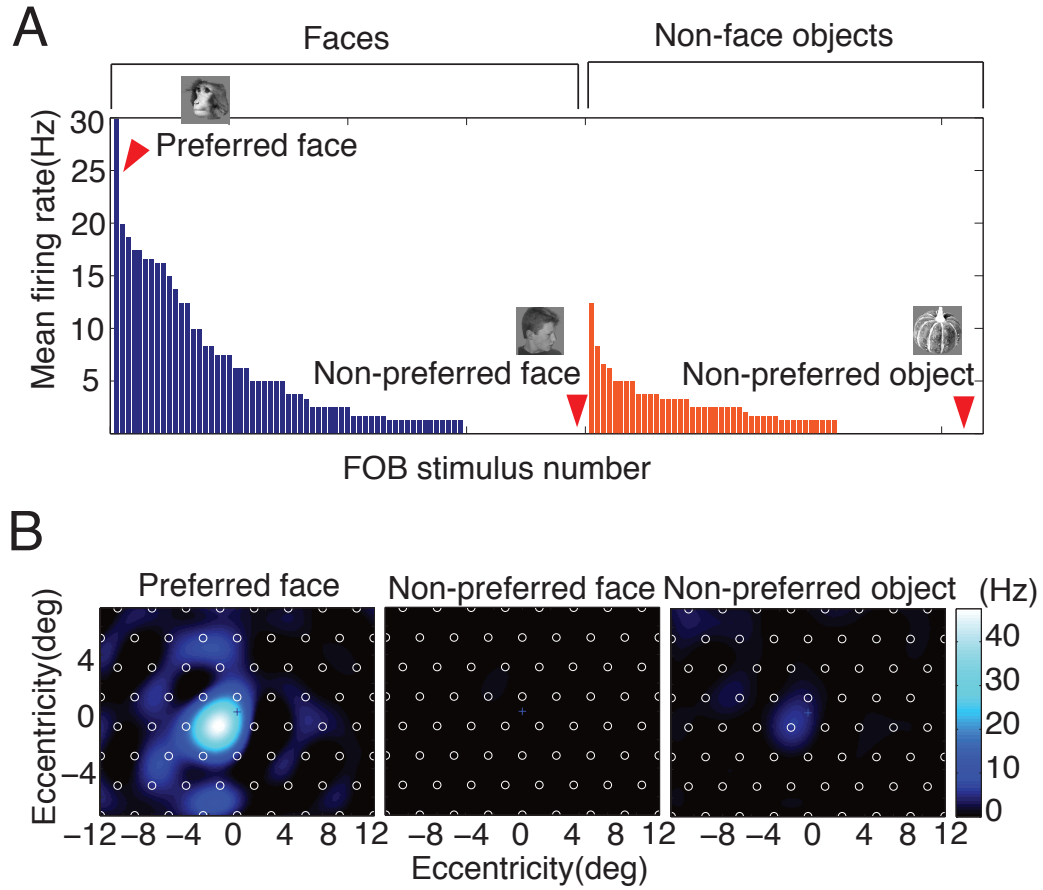


Figure 2.8: Selection of stimuli and stimulus design. (A) Mean response to 147 different visual stimuli of an example neuron. The most effective face stimulus, the least effective face stimulus and the least effective non-face object distractor are defined as preferred face, non-preferred face and non-preferred object. (B) The defined preferred face, non-preferred face and non-preferred object are presented at random location on the screen, in random order to map RFs.

were presented in point symmetric manner except in 1 distractor condition: 1 distractor could take any of the 8 presentation position surrounding the center preferred face. Overall, 1, 2, 4 and 8 distractors had 8, 4, 2 and 1 position variations, respectively. These stimuli were presented at the center of the measured RF. For the Prox/Ecc experiment, distractor number is fixed to 8, but distractor distance (horizontal / vertical separation between images) was varied from 0, 2 to 4 degrees of visual angle. Also, the stimulus position was changed either the RF center or 5 degrees away from the RF center. Note that we used heterogeneous, conspecific face distractors for the Prox/Ecc experiment. All of these combinatorial stimuli were presented for 400 ms with 200 ms inter-stimulus interval to ensure enough late response phase (see Chapter 3).

In the Cat/Num experiment, we also included stimulus condition where multiple preferred faces were presented instead of distractors (multiple preferred face condition). Other stimulus conditions we used were a large preferred face, a large non-preferred face and, large non-preferred object, 9 non-preferred faces and 9 non-preferred objects. These stimuli conditions are intended to test whether neurons are coding area or number of stimulus but the result from these stimuli will not be presented in this thesis. For control stimulus conditions, we used same distractor configuration without the preferred face (distractor only stimuli). These controls were needed to ensure that the impact by face distractors and object distractors were equally small, independent on the distractor numbers and / or distractor proximity. In total, the Cat/Num experiment contained 27 stimulus conditions. We presented the Cat/Num experiment for 8.5 minutes, resulting in approximately 30 repeats per stimulus, as shown in Equation 2.3. The Prox/Ecc experiment was presented for 5 minutes, resulting in approximately 30 repeats / stimulus, as shown in Equation 2.4⁷.

⁷The actual recorded trials could be less than these because we excluded trials when the subject was not fixating on the screen, and frame loss occurred during the presentation.

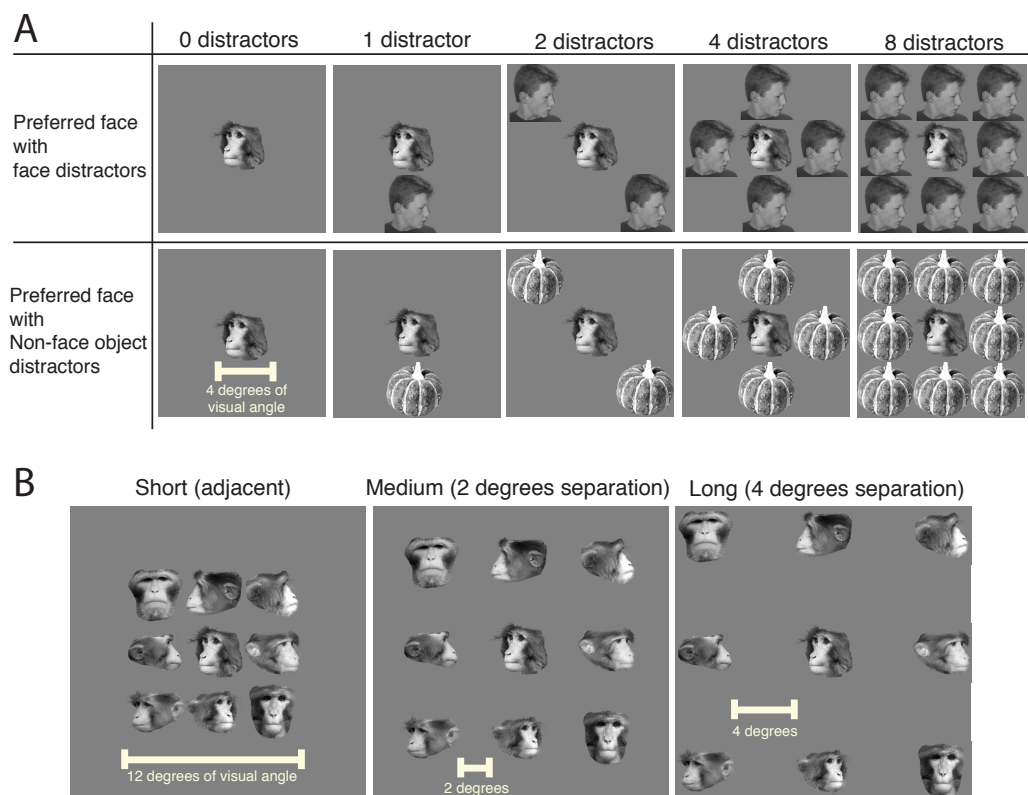


Figure 2.9: Design of multiple stimuli. (A) distractor number / category variation stimuli. Preferred face was presented at the center, and 1,2,4 or 8 distractors were presented in surround positions. 1,2 and 4 distractors were presented in point symmetry. (B) Prox/Ecc stimuli. Conspecific 8 different face distractors (i.e. the 8 least effective conspecific distractors) were presented with 0, 2 or 4 degrees separation. The stimuli were presented at either the center of measured RF or 5 degrees away from the center, toward peripheral visual field.

Fig. 2.10 shows the overall workflow of the experiments for a given single neuron. It took approximately 22.5 minutes per neuron to complete these stimuli (however when control experiments were presented also, the experiment took longer).

$$\frac{600[ms/image] \cdot 27 \cdot x[image]}{10^3[ms/sec] \cdot 60[sec/min]} = 8.5[min] \quad (2.3)$$

$$x \approx 31.5[repeats]$$

$$\frac{600[ms/image] \cdot 8 \cdot x \cdot 2[image]}{10^3[ms/sec] \cdot 60[sec/min]} = 6.5[min] \quad (2.4)$$

$$x \approx 31.2[repeats]$$

All of the stimuli shown in the Fig. 2.9 have the preferred face at the center position, and placed at the center of RF. This stimulus design can potentially make a confound effect of stimulus position: does a center stimulus acquire privilege to have more representational resource, more than response magnitude explained by the RF shape? Or, do face selective neurons have a single principle to allocate representational resource to each visual image only based on RF response magnitude at the position where the image was presented? To discriminate these two possibilities, we ran position control experiments on subtraction of the neurons: the preferred face is placed at the surrounding location and a distractor was placed at the center position. Using these stimuli, we studied if multiple visual stimuli could be represented, and if so, how the representation differed across neurons and time.

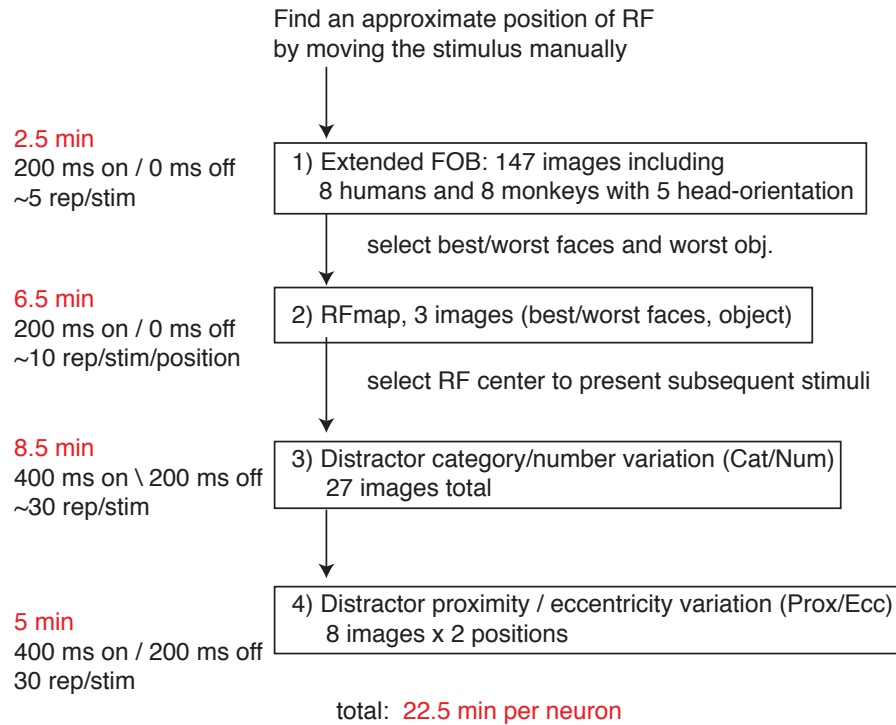


Figure 2.10: Experiment flow chart. Upon encountering a neuron, FOB stimulus was moved across the stimulus presentation screen to locate approximate position of the RF. Then, (1) FOB stimulus was presented at the putative center of the RF to characterize stimulus selectivity. Next, using 3 selected stimuli (preferred/non-preferred face and a non-preferred object), (2) RF was mapped. RF center was manually selected based on RF calculated online to present subsequent stimuli. (3) Cat/Num experiment was presented at the RF center, and (4) Prox/Ecc experiment was presented at the RF center and 5 degrees away from the RF center.

2.5 Targeting the face patch MF and AL

Three monkeys, two of *Macaca mulatta* (monkey Q, M) and one *Macaca fascicularis* (monkey J), were used for the recordings. MF recordings were conducted from monkey Q and J, whereas AL recordings were from monkey J and M. Prior to recording sessions, we needed to localize the target face patches. The monkeys were trained to keep fixation on a stimulus presentation screen inside a fMRI scanner, and face patches were defined based on BOLD activity to face and non-face object images. During the fMRI acquisition, MION (monocrystalline iron oxide nanoparticle) contrast agent was used to enhance signal-to-noise ratio. Acquired fMRI data was overlaid with anatomical MRI to be used in planning recording chamber placement.

To position the recording chamber, we used either Caret [156], OsiriX [157] or Planner [158] software to calculate the position and angle of the chamber. All three software packages display horizontal, vertical and sagittal section (anatomical sections) of the brain image to visualize 3D image of the brain. Therefore, we defined the angle of recording chamber within each section (e.g. rotate 15 degrees to lateral side within coronal plane, and rotate 5 degrees side). However, the common stereotaxic arm does not allow the user to make a specific angle within each anatomical section. Rather, it has only two degrees of freedom: tilt the stereotaxic arm and rotate the base of the stereotaxic arm. Thus, in order to implant the planned recording chamber angle using the stereotaxic arm, coordinates in the anatomical sections had to be converted into stereotaxic arm coordinate. For detailed calculations, see Appendix C.

Within the implanted recording chamber a craniotomy was made and a tungsten electrode (0.4 - 3 $M\Omega$ measured at 1000 Hz , FHC) was lowered through the dura guided by a custom-made metal guide tube and Crist grid

system [159]. The electrode was advanced by Narishige drive (MO-97A) for pre-calculated traveling distance until few millimeters before the target face patch. Then FOB stimulus was used to test neural selectivity to see if neurons were selective to faces. Once a recorded neuron was defined to be face selective, we started experiment summarized in Fig. 2.10. For other details of experimental procedures, see Appendix A.

3

Single neuron response properties and population average

3.1 Single neuron response and population average to Category/Number variation experiment

45 and 69 neurons were recorded from the middle face patch MF of two male adult rhesus macaques, *Macaca mulatta* (monkey Q) and *Macaca fascicularis* (monkey J), respectively. Of 114 recorded neurons, 111 and 108 neurons were analyzed for Cat/Num and Prox/Ecc experiment, respectively. Most of the analyses presented in this chapter are based on mean firing rates either across trials or neurons. For the population readout analyses (taking response pattern of the population into account) and computer modeling, see Chapter 4 and 5.

Fig. 3.1 shows an example neuronal response to the Cat/Num stimuli. The

neuron showed an initial peaked response and later plateau phase, which we defined as the early and the late phase, respectively. The borders of each phase were defined as follows. The early phase was the first 100 ms of the high amplitude response after the response onset, and the late phase was defined such that it contains approximately equal number of spikes as the early phase. Specifically, spikes across all the recorded neurons were accumulated for calculation, and we found 250 ms time window from 150 ms to 399 ms after the response onset contains approximately equal number of spikes as the early phase. Thereby we can (a) equalize the noise level (i.e. standard error of the mean is proportional to inverse square root of entry numbers) , and (b) calculate discriminability per spike in these regions, as shown in Chapter 4. The response onset was defined as the time when the mean firing rate crossed the threshold, which was 3 standard deviations of the spontaneous activity.

As face distractor number increased from 1, 2, 4, to 8, the neuron showed systematic response reduction both in the early and the late phase (Fig. 3.1, left). Two-way ANOVA (factor 1: distractor number, factor 2: distractor category) indicated that mean firing rate of the example neuron was reduced significantly both at the early phase ($F(4, 226) = 35.72, p = 4.04 \cdot 10^{-23} < 0.001$) and at the late phase ($F(4, 226) = 10.39, p = 9.63 \cdot 10^{-8} < 0.001$). Normalized mean firing rate (arbitrary unit) and standard error of the mean (SEM) with statistics are shown in Table 3.1.

For the detailed statistics of Tukey’s post-hoc pairwise significance test, see Appendix J. When non-face object distractors were used, however, the increasing number of object distractors also gave response reduction, but to a lesser degree (Fig. 3.1, right and Table 3.1). Two-way ANOVA confirmed that the mean firing rates with object distractors were significantly larger than

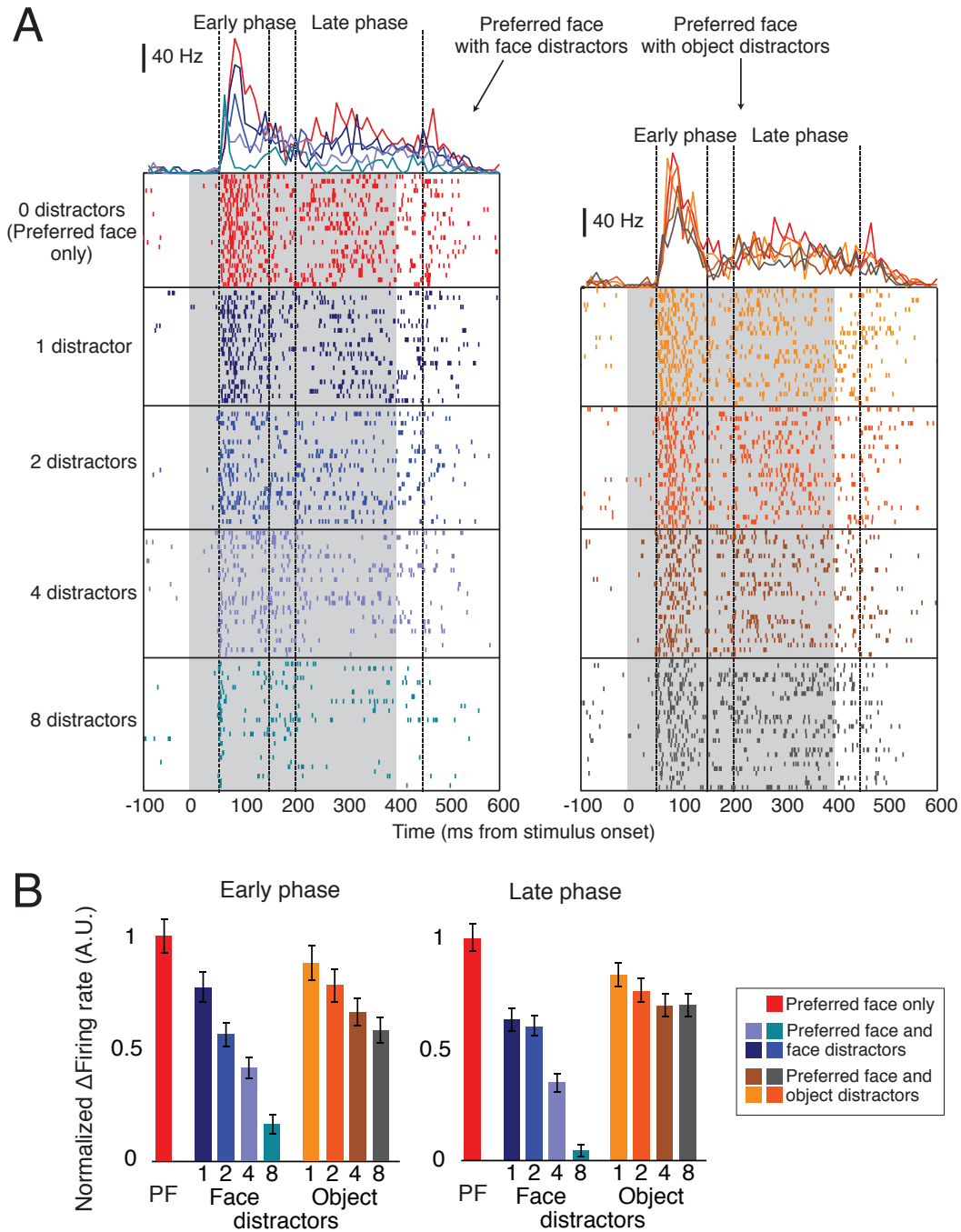


Figure 3.1: Example neuronal response to the Cat/Num stimuli. (A) The raster plot shows the occurrences of the spike in each trial for face (left) and object (right) distractors, and the average of the raster plots is shown above. Abscissa is time in millisecond (ms), and raster plots show each trial response. Gray shadow is indicating the stimulus presentation duration, 400 ms. Dotted lines indicate early and late phase, defined based on response latency. (B) Normalized mean firing rates calculated within the time windows of the early (left) or the late (right) phase.

Table 3.1: Cat/Num experiment, MF single cell average response

	Early		Late	
	Face	Obj	Face,	Obj
0 dist.	1 ± 0.0315	1 ± 0.0315	1 ± 0.0641	1 ± 0.0641
1 dist.	0.793 ± 0.0498	0.890 ± 0.0347	0.698 ± 0.0473	0.857 ± 0.0590
2 dist.	0.598 ± 0.0673	0.792 ± 0.0289	0.665 ± 0.0746	0.804 ± 0.0552
4 dist.	0.460 ± 0.0555	0.691 ± 0.0456	0.458 ± 0.0776	0.751 ± 0.0741
8 dist.	0.228 ± 0.0362	0.611 ± 0.0388	0.202 ± 0.0465	0.749 ± 0.0913

the firing rates with face distractors, both at the early phase ($F(1, 226) = 47.69, p = 5.00 \cdot 10^{-11} < 0.001$) and at the late phase ($F(1, 226) = 29.37, p = 1.53 \cdot 10^{-7} < 0.001$). Especially at the late phase, responses were robust against the object distractor number increase. When one-way ANOVA was used to test effect of object distractor number on mean firing rate (without considering face distractor group), p-value did not reach significance ($F(3, 102) = 0.49, p = 0.692 > 0.05$). Tukey’s post-hoc test also showed no significance between any pairs in the object distractor conditions at the late phase (see Appendix J).

Neurons recorded from MF showed large cell-to-cell difference in response latencies (mean, $\mu = 90.3$, standard deviation, $\sigma = 49.2$ ms after the stimulus onset). Thus, averaging across neurons involved a risk of mixing different phases of responses resulting in obscuring the result. Thus, we subtracted response latency from each neuron to align the response onset before averaging across neurons. Note that we preserved the latency difference, if any, across different stimulus conditions. In other words, we defined one response latency per neuron, and subtracted from all the responses recorded from that neuron. The response latencies were calculated based on an average response to the preferred face. The resulting latency-subtracted average firing rate showed a qualitatively similar response pattern as population average without latency subtraction (See Appendix F), but with a sharper initial rise (Fig. 3.2). Here-

after I use the term “population average” to mean the latency-subtracted average firing rate unless otherwise noted, but the result of all other analyses were qualitatively similar for population average without latency subtraction. The population average may or may not be normalized such that the maximum firing rate equals 1. Spontaneous activity was subtracted before averaging, thus the baseline equals 0. Overall, the average firing rate across 111 neurons showed a qualitatively similar response pattern as the single cell example shown in Fig. 3.1. Increasing the number of distractors systematically reduced the response, and face distractors reduced responses more than object distractors did (see Table 3.2 for mean firing rates and SEM).

Table 3.2: Cat/Num experiment, MF population average response

	Early		Late	
	Face	Obj	Face,	Obj
0 dist.	1 ± 0.0453	1 ± 0.0453	1 ± 0.0603	1 ± 0.0603
1 dist.	0.722 ± 0.0425	0.826 ± 0.0444	0.738 ± 0.0508	0.795 ± 0.0471
2 dist.	0.504 ± 0.0387	0.692 ± 0.0406	0.633 ± 0.0565	0.783 ± 0.0546
4 dist.	0.389 ± 0.0359	0.575 ± 0.0397	0.571 ± 0.0584	0.738 ± 0.0565
8 dist.	0.306 ± 0.0340	0.496 ± 0.0397	0.443 ± 0.0603	0.766 ± 0.0640

We quantified this effect by two-way ANOVA (factor 1: distractor category, factor 2: distractor number). In both the early and the late phase, face distractors significantly suppressed the firing rate compared to object distractors (early phase: $F(1, 880) = 40.06, p = 3.92 \cdot 10^{-10} < 0.001$, late phase: $F(1, 880) = 26.87, p = 2.70 \cdot 10^{-7} < 0.001$). Reduction by the number of distractors was also significant (early phase: $F(3, 880) = 41.06, p < 10^{-25} < 0.001$, late phase: $F(3, 880) = 6.44, p = 2.60 \cdot 10^{-4} < 0.001$)¹. In order to find which pair-

¹For the completeness: synergetic interaction between category and numbers did not reach significance in neither phase (early: $F(3, 880) = 0.58, p = 0.631$, late: $F(3, 880) = 2.17, p = 0.0898$).

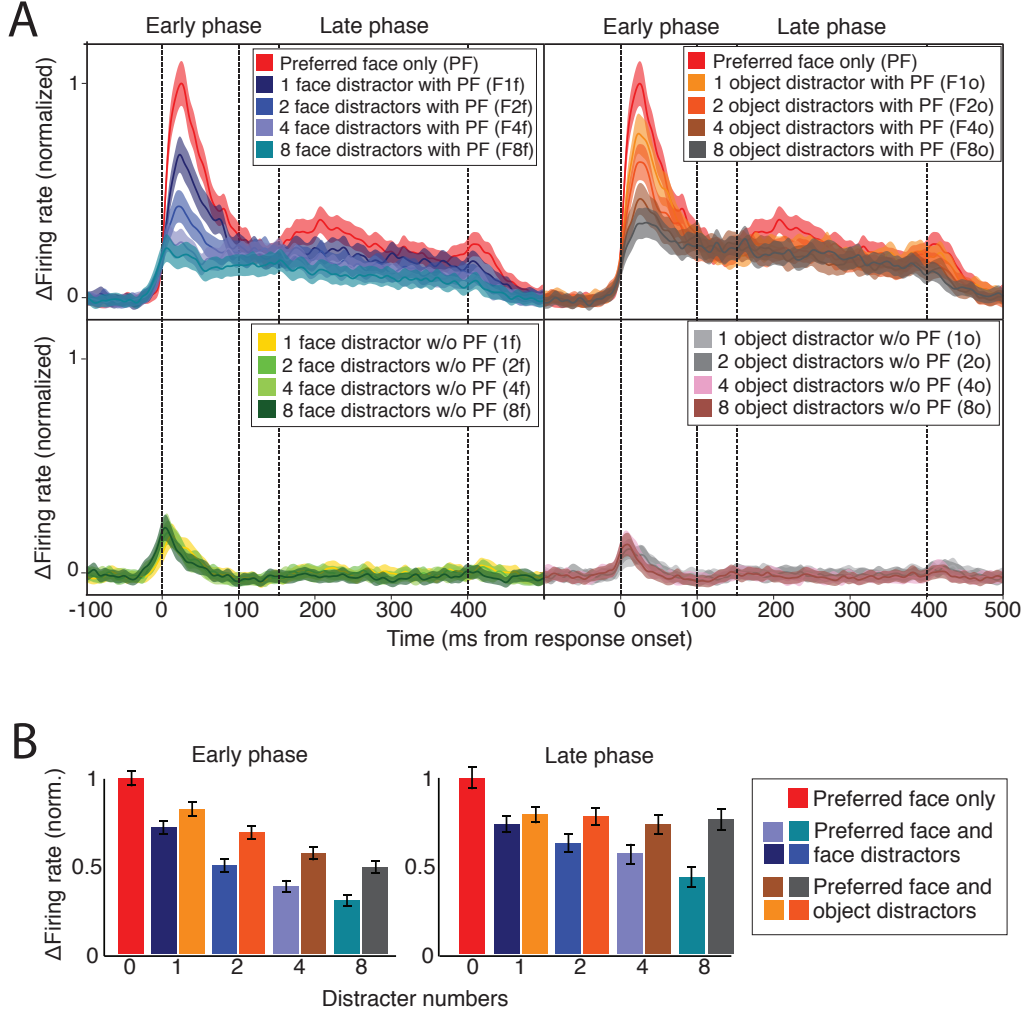


Figure 3.2: Population average response to the Cat/Num stimuli. (A) Latency subtracted population average (N=111). PSTHs show the time course of responses to the preferred face with/without distractors (top) and distractors without preferred face (bottom). (B) Mean firing rates of the population average at the early and the late phase. PF: preferred face, F1f, F2f, F4f, F8f: 1,2,4 or 8 face distractors with preferred face, F1o, F2o, F4o, F8o: 1,2,4 or 8 object distractors with preferred face, 1f, 2f, 4f, 8f: 1,2,4 or 8 face distractors without preferred face, 1o, 2o, 4o, 8o: 1,2,4 or 8 object distractors without preferred face.

wise significance, we conducted Tukey’s post-hoc test (Fig. 3.3). In the early phase, most of the pairs showed significance, indicating systematic response reduction by increasing number of distractors. In the late phase, however, the object distractors and preferred face condition (Fig. 3.3B, right panel) showed significance between zero distractors (preferred face only) and 2, 4, 8 distractors conditions but pairs across 1,2,4 or 8 distractors did not reach significance, confirming our observation in Fig. 3.2: different numbers of the object distractors do not affect firing rates as much as face distractors.. For the detailed statistics of Tukey’s post-hoc test, see Appendix J. Thus, in the Cat/Num experiment, the mean firing rate of single neurons was reduced with increasing number of distractors, and/or with object distractors rather than face distractors.

We found a categorical difference of the distractor in suppressing the response to the preferred face. Namely, face distractors suppress the response of face selective neurons more than object distractors. One trivial, uninteresting explanation would be that object distractors themselves are driving the neurons more strongly than face distractors: that’s why the preferred face with the object distractors gave larger responses. However, this possibility was excluded. As shown in Fig. 3.2A (bottom), the response to the face distractors and object distractors alone (i.e. no preferred face) elicited quite comparable and weak response at the early phase. At the late phase, responses were almost at the baseline² (i.e. spontaneous activity) and difference was not significant

²At the early phase, the responses to face distractors themselves were significantly different from the responses to object distractors (with Two-way ANOVA, early phase: $F(1, 883) = 10.3, p = 0.0014 < 0.05$, late phase: $F(1, 883) = 0.41, p = 0.521 > 0.05$). Although the statistics reached the significance, however, it cannot hamper our argument: if the observed larger response reduction due to the face distractor is because of smaller drive provided by face distractor themselves, the face distractor need to elicit larger response than object distractors. However, the opposite result was shown in here: object distractors elicited smaller response. In any case, the difference between responses to distractors without preferred face was too small to explain the suppression by intra-category (i.e. face) distractor and by increasing distractor numbers.

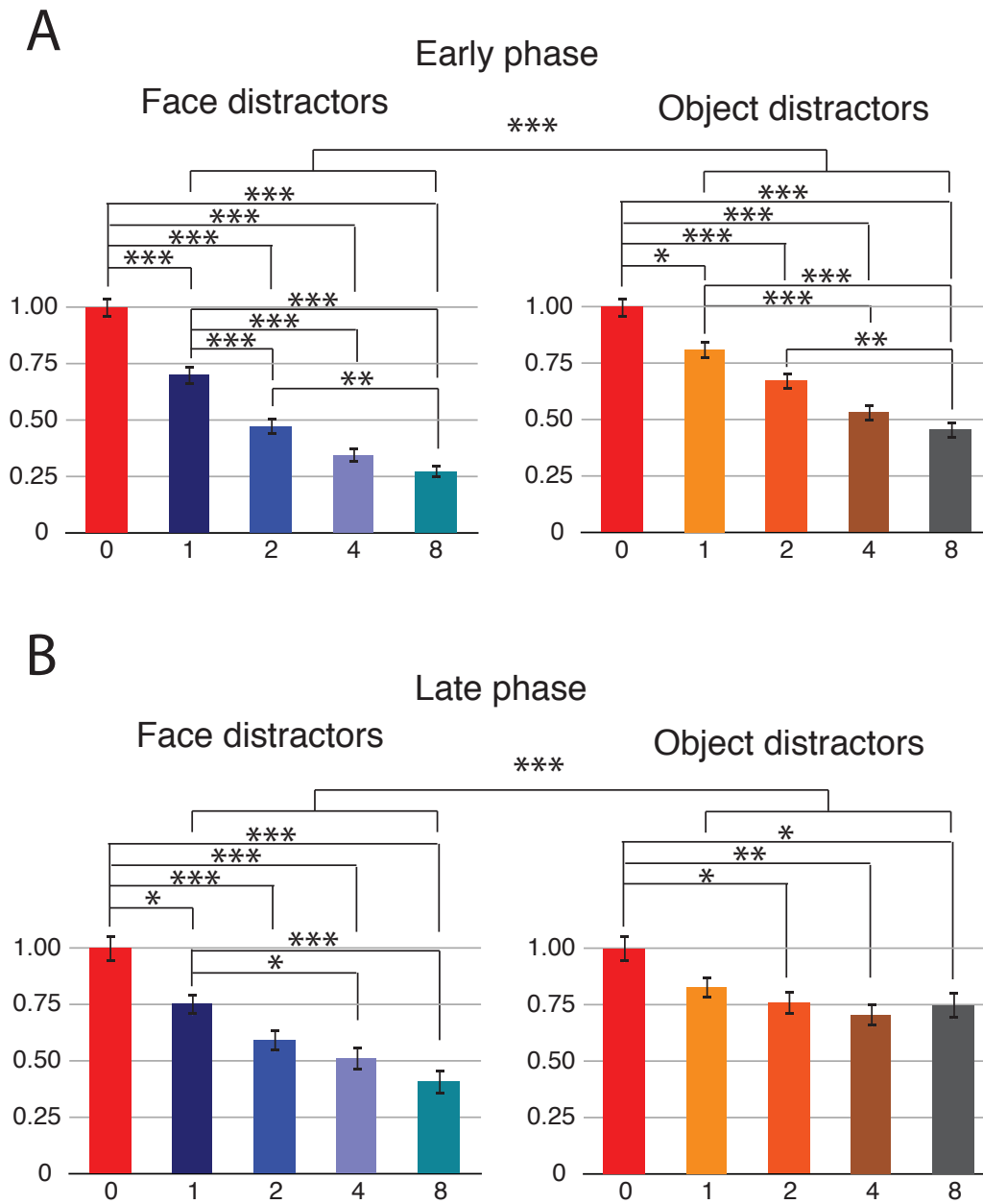


Figure 3.3: Significance test from ANOVA and Tukey's post-hoc test. (A) Early phase, face distractors and preferred face conditions (left) and object distractors and preferred face conditions (right). Bar heights were normalized to highlight the proportion of reduction. (B) Same as (A) but with the late phase data.

either ($F(1, 883) = 0.41, p = 0.521 > 0.05$). Therefore, the distractor category-dependent suppression cannot be ascribed to the difference in the magnitude of the response distractors elicited to neurons. See Table 3.3 for mean firing rates and SEM.

Table 3.3: Cat/Num distractor-only control, MF population average response

	Early		Late	
	Face	Obj	Face,	Obj
1 dist.	0.101 ± 0.0169	0.0524 ± 0.0125	0.00369 ± 0.0198	-0.0133 ± 0.0176
2 dist.	0.0927 ± 0.0182	0.0497 ± 0.0150	-0.0146 ± 0.0215	-0.0208 ± 0.0201
4 dist.	0.0862 ± 0.0200	0.0470 ± 0.0176	-0.0372 ± 0.0204	-0.0519 ± 0.0195
8 dist.	0.0846 ± 0.0205	0.0333 ± 0.0186	-0.0601 ± 0.0221	-0.0548 ± 0.0205

As shown in Fig. 3.2A, difference between responses elicited by face and object distractors were very small at the early phase, and almost no difference at the late phase. No significance was found across different distractor numbers at the early phase (ANOVA, $F(3, 883) = 0.37, p = 0.7753 > 0.05$) but weak significance was found at the late phase ($F(3, 883) = 2.83, p = 0.0376 < 0.05$).

Assuming that response to distractors themselves (i.e. without the preferred face) were sufficiently weak compared to the response to stimulus condition containing the preferred face³, we compared the impact of face and object distractor by normalizing the mean firing rates and fitting a rational function to them:

$$R = \frac{1}{1 + \alpha x} \quad (3.1)$$

³This assumption was verified with ANOVA: the responses to 1,2,4 or 8 distractors (either faces or objects) with or without the preferred face were significantly different indicated by 1-way ANOVA $F(1, 1774) = 855.37, p = 8.59 \cdot 10^{-154} < 0.001$ at the early phase and $F(1, 1774) = 1156.91, p = 1.17 \cdot 10^{-195} < 0.001$ at the late phase. Especially at the late phase, the response to distractors was almost flat (or slightly suppressive for some neuron).

In Equation 3.1, α is a free parameter to fit and x is distractor number (0, 1, 2, 4, 8). Recorded neural responses were averaged at the early or at the late phase, and normalized such that the response to the preferred face was equal to one. Within this configuration, if α is equal to zero, dependency on distractor number disappears and a neuron become a max neuron. If α is equal to one instead, the neuron is taking an average of the inputs provided by the preferred face and distractor(s), given that the response to distractor themselves are sufficiently smaller than 1. Fig. 3.4 shows that at the early phase (panel A) most of the neuron had the α value in between 0 and 1, and α values were smaller for face distractor conditions (i.e. most of the values are under the diagonal line of unity), confirming that face distractors had larger impact on reducing the mean firing rate. At the late phase, α values shifted toward zero, indicating that at the late phase more neurons behaved similar to the max neuron. Few neurons showed max-like response in both the early and the late phase (Fig. 3.4, black points near the origin). We investigate this finding and the underlying mechanism further in Chapter 5.

3.2 Single neuron response and population average to Proximity/Eccentricity variation experiment

Fig. 3.5 shows the raster plot and average firing rate of an example neuron responding to Prox/Ecc stimulus. As distractor distance became shorter, the firing rate of the example neuron was reduced. One-way ANOVA testing differences in mean firing rates in the 4 stimulus conditions showed decrease in both the early phase ($F(3, 162) = 40.26, p = 1.70 \cdot 10^{-19} < 0.001$) and the late

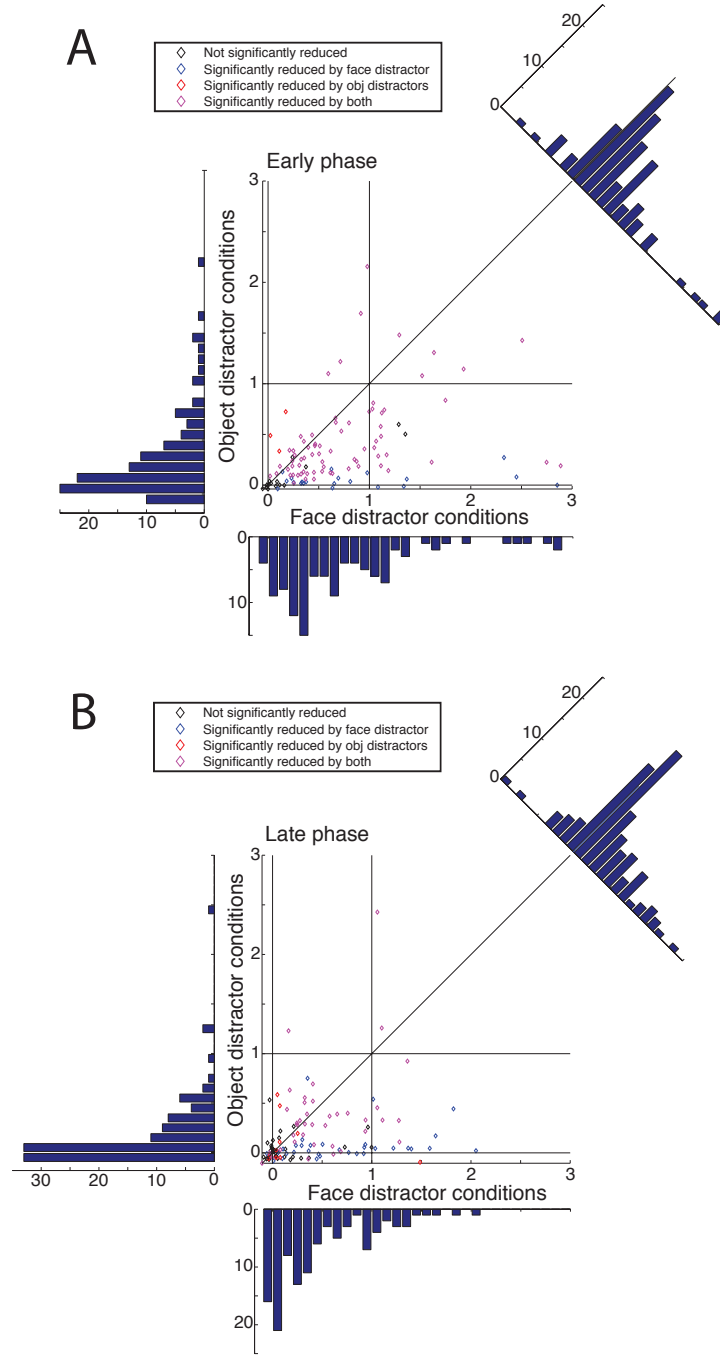


Figure 3.4: Coefficient from rational function fitting to compare distractor impact. Ordinate and abscissa represent object and face distractor conditions, respectively. α value from Equation 3.1 was plotted with color code indicating the statistical significance. Panel (A) and (B) show the early and the late phase, respectively. Histograms show marginal distributions.

phase, although the latter did not reach significance in this particular example neuron ($F(3, 162) = 1.08, p = 0.3596 > 0.05$). See Table 3.4 for mean firing rates and SEM.

Table 3.4: Prox/Ecc experiment, MF single cell average response

	Early		Late	
	Center	Periphery	Center,	Periphery
0 dist.	1 ± 0.100	1 ± 0.161	1 ± 0.138	1 ± 0.122
L dist.	0.378 ± 0.0381	0.415 ± 0.0881	0.899 ± 0.0737	0.451 ± 0.0570
M dist.	0.331 ± 0.0429	0.435 ± 0.0661	0.776 ± 0.0713	0.332 ± 0.0357
S dist.	0.163 ± 0.0234	0.170 ± 0.0474	0.805 ± 0.0736	0.419 ± 0.0528

When the stimuli were presented at the periphery of RF, the distractors also suppressed neural responses, shown by one-way ANOVA (The early phase: $F(3, 153) = 11.53, p = 7.37 \cdot 10^{-7} < 0.001$, The late phase: $F(3, 153) = 14.84, p = 1.57 \cdot 10^{-8} < 0.05$). The mean firing rate between stimulus groups including distractors with a preferred face (the preferred face with large, medium or short distance distractors) and without a preferred face (large, medium or short distractors without the preferred face) was also significant at RF center (The early phase: $F(1, 287) = 64.45, p = 2.56 \cdot 10^{-14} < 0.001$, the late phase: $F(1, 287) = 152.66, p = 2.08 \cdot 10^{-28} < 0.05$). However, at RF periphery, the presence of a preferred face did not show significantly different firing rates (The early phase: $F(1, 281) = 1.88, p = 0.172 > 0.05$, the late phase: $F(1, 281) = 1.82, p = 0.179 > 0.05$). This was because at RF periphery, addition of distractors suppressed neural response enough so that the responses were indistinguishable from the responses to distractors themselves. Additionally, distractors with decreasing distance did not elicit systematic response reduction as seen in RF center presentation. For Tukey's post-hoc test showing

pairwise significance, see Appendix J. See Table 3.4 for mean firing rates and SEM.

Population average firing rate (Fig. 3.6) of the 106 recorded neurons⁴ showed qualitatively same pattern in the early and the late phase. Similar to the Cat/Num experiment, we subtracted the response onset latency to average responses across recorded neurons. As shown in Fig. 3.6, distractors with decreasing distances reduced the mean firing rate significantly at the early phase ($F(3, 428) = 37.6, p = 1.38 \cdot 10^{-21} < 0.001$) and at the late phase ($F(3, 428) = 27.97, p = 1.55 \cdot 10^{-16} < 0.001$). Distractors also reduced mean firing rates when stimuli were presented at the RF periphery. Both at the early phase ($F(3, 428) = 20.3, p = 2.57 \cdot 10^{-12} < 0.001$) and at the late phase ($F(3, 428) = 24.08, p = 2.04 \cdot 10^{-14} < 0.001$) ANOVA showed significance response reduction. See Table 3.5 for mean firing rates and SEM.

Table 3.5: Prox/Ecc experiment, MF population average response

	Early		Late	
	Center	Periphery	Center,	Periphery
0 dist.	1 ± 0.0500	1 ± 0.0924	1 ± 0.100	1 ± 0.103
L dist.	0.596 ± 0.0502	0.386 ± 0.0579	0.601 ± 0.0529	0.282 ± 0.0654
M dist.	0.497 ± 0.0481	0.346 ± 0.0556	0.476 ± 0.0504	0.199 ± 0.0640
S dist.	0.323 ± 0.0368	0.381 ± 0.0664	0.319 ± 0.0492	0.227 ± 0.0730

Consistent with the observation in the single neuron, one difference we found between RF center and RF periphery condition was the significance between different distractor proximity. For example, the preferred face and short distance distractor condition (PF+SD) gave significantly smaller mean firing rate compared to the preferred face and large distance distractor condition

⁴The reason why we have fewer recorded neuron in the Prox/Ecc experiment is simply because this experiment was conducted at the end of the entire recording stream for one neuron: longer recording time entails higher risk of losing a neuron (see Chapter 2).

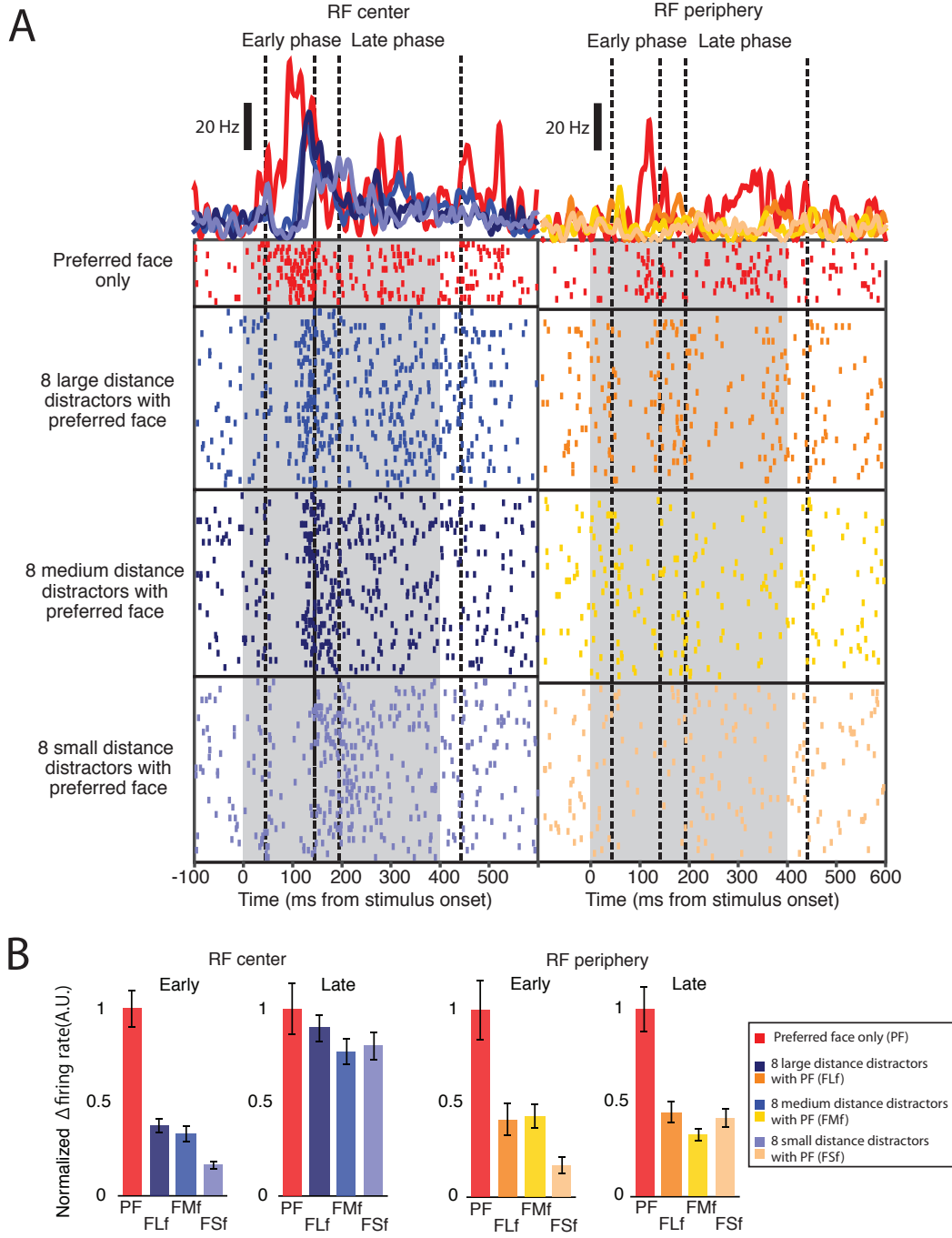


Figure 3.5: An example neural response to the Prox/Ecc experiment. (A, left) Stimulus presentation at the RF center. (A, right) Stimulus presentation at RF periphery. Raster plots show each trial response. Gray shadow indicates stimulus presentation duration (400 ms with 200 ms inter-stimulus interval). Top PSTH indicates mean firing rate in 10 ms time bin. Dotted lines indicate early and late phase, defined based on response latency. (B) Normalized mean firing rates calculated within the time windows of the early (left) or the late (right) phase.

(PF+LD) when stimuli were presented at RF center. In the RF periphery presentations, however, this condition did not show significance. In other words, at the RF periphery, neurons reduce their response in presence of distractors, but did not change response magnitude with distractor proximity. For details of Tukey’s post-hoc test result, see Appendix J.

3.3 Response dependency on receptive field size

One might expect that the suppressive effect caused by distractors depends on RF size. For example, a neuron with a smaller RF excludes distractors out of RF to have robust response to the preferred face regardless of the presence of distractors. However, the results so far looked independent of neuronal RF size: both large RF neurons and small RF neurons showed similar response reductions in both Cat/Num and Prox/Ecc experiment (Fig. 3.7A,B). We quantified RF size by calculating RF index (see Appendix A) and calculated correlation coefficient between the index and response reduction magnitude by face distractor numbers ($r = -0.120$ (early phase) / 0.0685 (late phase)), by object distractor numbers ($r = -0.119$ (early phase) / 0.0685 (late phase)), or by face/object category difference ($r = 0.112$ (early phase) / 0.094 (late phase)) but correlation did not reach significance (p-values are 0.237 (early) / 0.546 (late), 0.244 (early) / 0.546 (late), 0.253 (early) / 0.409 (late), respectively). Thus, response reduction by distractors are consistent across recorded neurons independent of RF size.

However, when multiple preferred faces were presented, the response was highly predicted by the RF size (Fig. 3.7C). The correlation between RF size index and the response to multiple preferred faces were highly significant ($r = 0.421, p = 1.58 \cdot 10^{-5} < 0.001$). When a RF can encompass all the preferred face presented, neural response stayed high or even increased with increasing

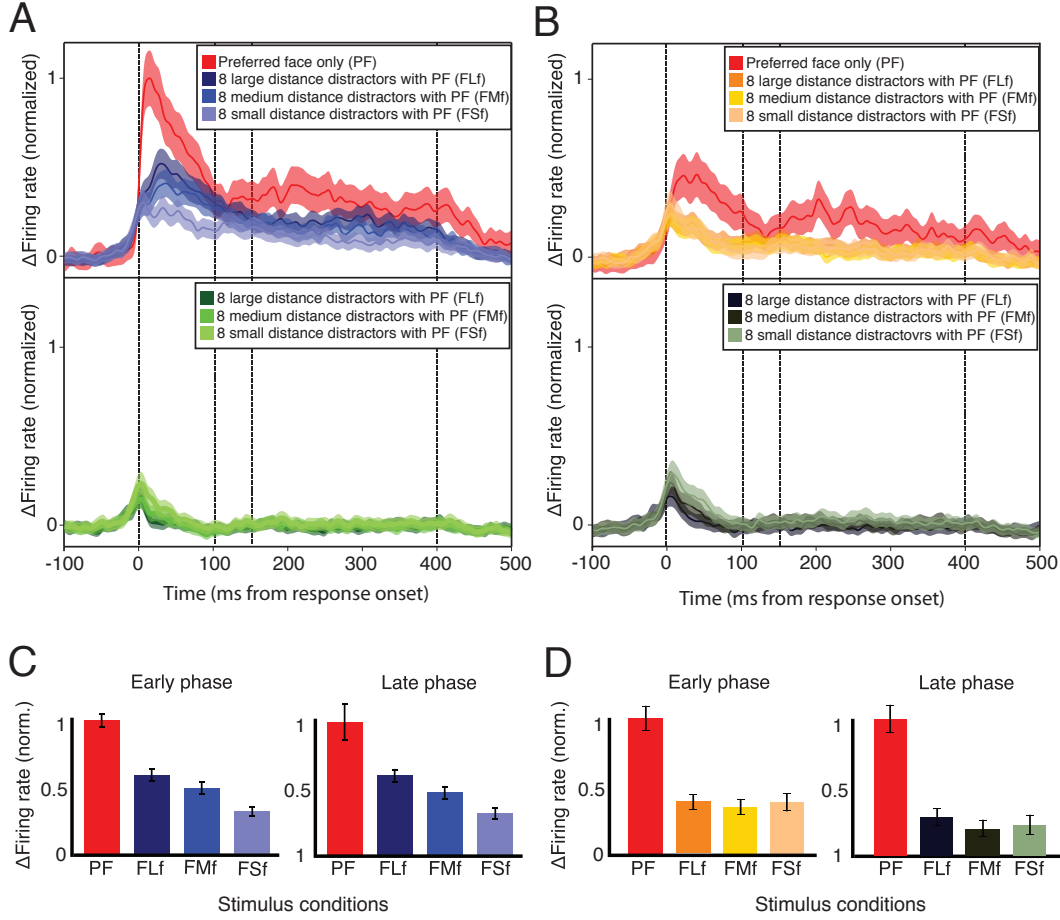


Figure 3.6: Population average response to the Prox/Ecc experiment. (A,C) Stimulus presentation at RF center. (A) Latency subtracted population average (N=108). PSTHs show the time course of responses to the preferred face with or without distractors (top) and distractors without preferred face (bottom). (C) Mean firing rates of the population average at the early and the late phase. PF: preferred face, FLf, FMf, FSf: preferred face with 8 large, medium or small distance distractors. Lf, Mf, Sf: 8 large, medium or small distance distractors without preferred face. (B,D) same as A and C, except that presentation location was RF periphery.

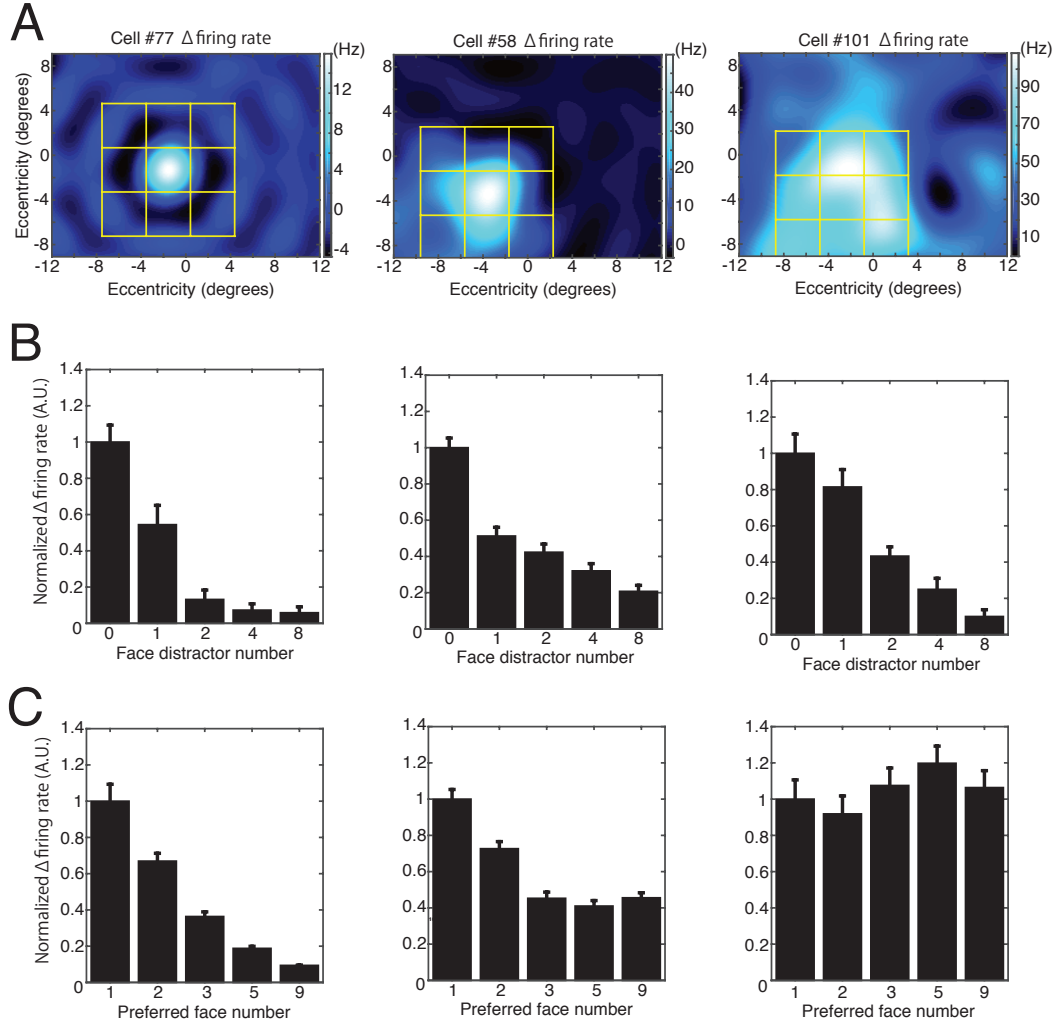


Figure 3.7: Response reduction dependency on RF size. Three example cells with (from left to right column) small, medium and large RFs are shown. (A) RFs of the three example neurons. Nine stimulus presentation locations are highlighted with yellow lines. (B) Responses to a preferred face (at the center) with 0,1,2,4 or 8 face distractors (at the surround locations), shown in normalized and spontaneous activity-subtracted firing rates. (C) Responses to multiple preferred faces, shown in normalized and spontaneous activity-subtracted firing rates.

number of preferred face (though the magnitude of increase was much smaller than a sum of all inputs). In contrast, however, if a RF is too small to encompass all of the preferred faces, the neural response was reduced, as if the preferred faces were acting like a non-preferred, distractor face. Thus, RF field size affects a neuronal integration of preferred faces, but not an integration of distracters.

3.4 Response reduction was not dependent on preferred face position

Finally, we compared the results from Cat/Num experiment to position control experiment. Fig. 3.8 shows the average firing rate of the 15 neurons from which the control experiments were recorded. The average firing rate in the position control experiment showed smaller firing rate magnitude, as expected from the fact that the preferred face was presented at the periphery of the RF, which is by definition less effective than the center. Similarly, responses to distractors without the preferred face showed larger firing rates than the Cat/Num experiment, also expected from the fact that a distractor occupied the RF center, the most effective visual field driving the neuron. See Table 3.6, 3.7, 3.8 and 3.9 for mean firing rates and SEM (note that firing rates are not normalized).

Nonetheless, average of the position control experiment showed response reduction similar as the Cat/Num experiment at the early phase (Two-way ANOVA, factor1: distractor number, factor2: Cat/Num vs. position control experiment, Cat/Num: $F(4, 129) = 7.6, p = 1.58 \cdot 10^{-5} < 0.001$ position control experiment: $F(4, 129) = 2.54, p = 0.043 < 0.05$), although late phase did not reached the significance (Cat/Num experiment: $F(4, 129) = 1.92, p =$

Table 3.6: Cat/Num experiment, average of 15 representative neurons

	Early		Late	
	Face	Obj	Face,	Obj
0 dist.	46.5 ± 1.73	46.5 ± 1.73	22.9 ± 0.325	22.9 ± 0.325
1 dist.	32.3 ± 1.10	39.3 ± 1.32	14.4 ± 0.182	14.9 ± 0.178
2 dist.	23.7 ± 0.709	29.6 ± 1.36	12.4 ± 0.232	15.4 ± 0.167
4 dist.	16.4 ± 0.501	23.1 ± 0.697	12.4 ± 0.280	15.0 ± 0.234
8 dist.	9.44 ± 0.440	19.1 ± 0.517	10.7 ± 0.153	14.1 ± 0.154

Table 3.7: Position control, average of 15 representative neurons

	Early		Late	
	Face	Obj	Face,	Obj
0 dist.	27.4 ± 1.20	27.4 ± 1.20	13.6 ± 0.175	13.6 ± 0.175
1 dist.	18.0 ± 1.00	20.0 ± 1.08	9.09 ± 0.169	8.08 ± 0.183
2 dist.	15.1 ± 0.807	16.9 ± 0.951	6.51 ± 0.148	8.21 ± 0.180
4 dist.	11.0 ± 0.807	13.4 ± 0.900	6.70 ± 0.0965	6.32 ± 0.167
8 dist.	7.06 ± 0.710	12.8 ± 0.821	4.31 ± 0.182	8.73 ± 0.211

Table 3.8: Cat/Num distractor-only control, average of 15 representative neurons

	Early		Late	
	Face	Obj	Face,	Obj
1 dist.	2.45 ± 0.368	0.794 ± 0.262	0.154 ± 0.0763	-0.974 ± 0.0879
2 dist.	2.31 ± 0.569	1.09 ± 0.317	-0.564 ± 0.0636	-0.581 ± 0.0623
4 dist.	1.93 ± 0.616	0.273 ± 0.350	-1.20 ± 0.0741	-0.258 ± 0.0879
8 dist.	1.85 ± 0.162	-0.217 ± 0.106	-0.735 ± 0.0618	0.549 ± 0.105

Table 3.9: Position control distractor-only, average of 15 representative neurons

	Early		Late	
	Face	Obj	Face,	Obj
1 dist.	5.26 ± 0.910	3.47 ± 0.796	5.26 ± 0.910	3.48 ± 0.796
2 dist.	3.59 ± 0.834	2.69 ± 0.738	3.59 ± 0.834	3.47 ± 0.796
4 dist.	2.98 ± 0.693	1.50 ± 0.639	2.98 ± 0.693	2.69 ± 0.738
8 dist.	3.99 ± 0.707	-0.0594 ± 0.418	43.99 ± 0.707	1.50 ± 0.639

0.111>0.05, position control experiment: $F(4, 129) = 1.31, p = 0.271 > 0.05$). No difference in mean values was found⁵ in the Cat/Num experiment and position control experiment, both in the early ($F(1, 264) \sim 0, p = 0.975 > 0.05$) and the late ($F(1, 264) = 0.57, p = 0.451 > 0.05$) phase.

3.5 Summary

To summarize, we found neurons reduced their firing rates to its preferred face, with increasing number of distractors and with face distractors, more than object distractors. These effects are independent on neural RF size. The only response we observed that depended on RF size was the response to multiple preferred faces, where faces outside of the RF cause response suppression similar to face distractors. The neuronal firing rates were also reduced with closer distractor proximity at RF center, but at RF periphery, distractors did not cause systematic response reduction with shortening distractor distance.

⁵In order to equalize the difference in the peak magnitude, both data in Cat/Num and position control experiment were normalized such that population average response to the preferred face is equal to 1.

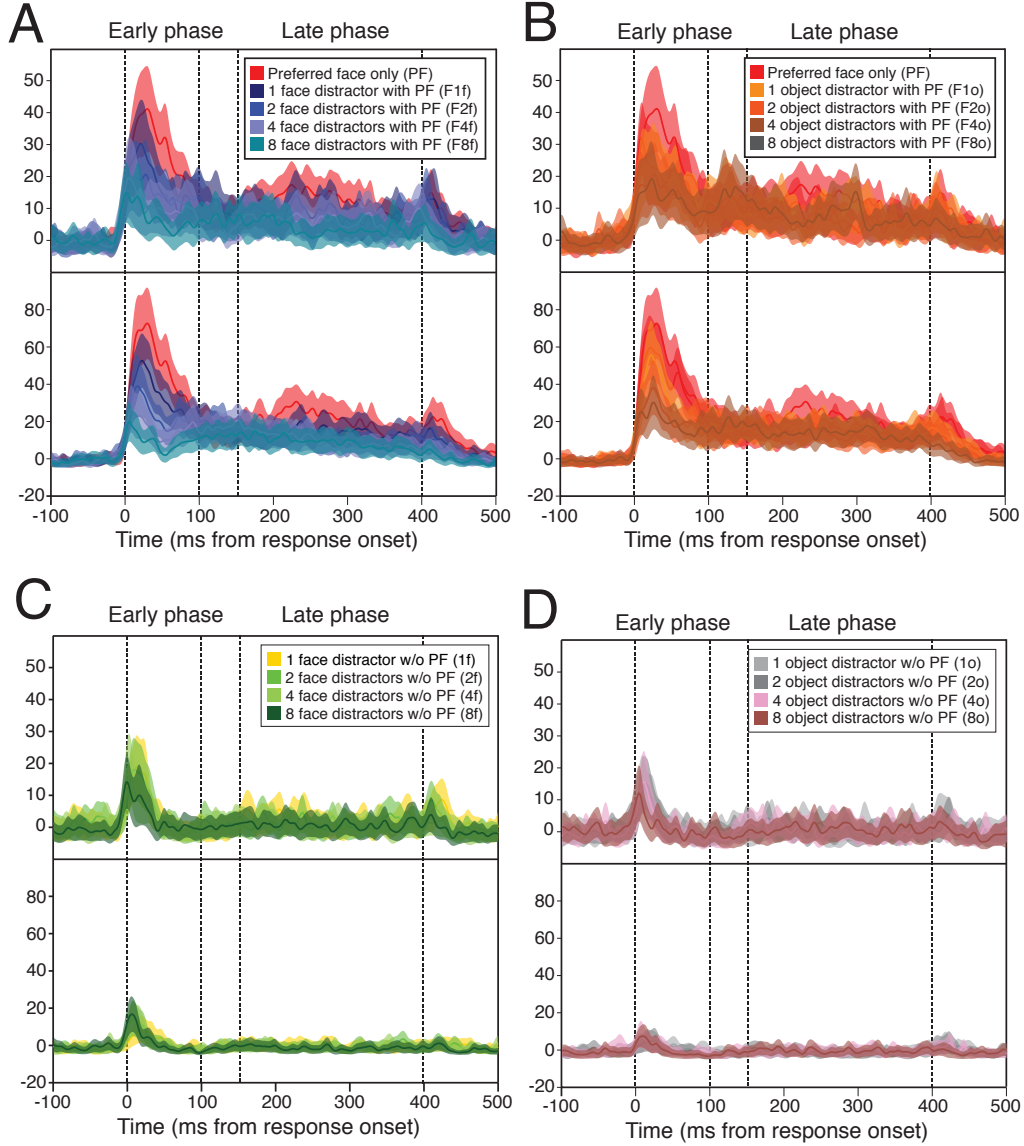


Figure 3.8: Position control experiment. Each panel shows latency subtracted population average (N=15) of position control experiment (above) with the preferred face presented at the peripheral RF, and main experiment (below) with the preferred face presented at the center RF. Note that above and below abscissa have different scale, but the scales are fixed across panels. (A) The preferred face and face distractor conditions. (B) The preferred face and object distractor conditions. (C) Face distractor without preferred face. (D) Object distractor without preferred face.

4

Population readout

4.1 Representation in the pattern of neural responses

We observed that with increasing number of distractors, with face distractors rather than object distractors, and with closer distractors the mean neuronal firing rate was reduced. This result might indicate that the neural representation of the preferred face was degraded by these distractors. However, while average firing rate is a simple and easy way to visualize the data, it can potentially obscure the representation within the response pattern across the recorded neurons. In other words, small fluctuation could create a pattern of activity in the neural population to encode the stimulus from the outside world, even if such a pattern could be averaged out. For example, Stokes et al. showed that the response pattern of neural population could encode information, even under a very small magnitude change in average firing rate [160].

The reason why the averaging operation can compromise the actual stimulus representation can be seen in Fig. 3.2. In the top two panels, “preferred face

with 2 face distractors” and “preferred face with 4 object distractors” show approximately same average response. However, these two stimulus conditions could be represented very differently in the representation space spanned by neural population. Fig. 4.1 provides an explanation. The Top panel shows a representation space spanned by two hypothetical face selective neurons, responding to face distractor and/or object distractor, in addition to a preferred face. Each dot is showing two neurons’ responses to a stimulus combination. Two entries in the parenthesis indicate whether the face distractor and object distractor is present (“1”) or absent (“0”), respectively. For example, B(1, 0) means face distractor is present, and object distractor is absent. Averaging corresponds to the following operation: project each dot to the diagonal line of unity ($x = y$. See Appendix H for explanation of why diagonal projection corresponds to averaging). The result is shown in the bottom left panel of Fig. 4.1. The average correctly captures the large activity difference between no distractor presence (D(0, 0), preferred face without any distractor) and both distractor presence (A(1, 1)) condition. However, the condition B(1, 0) and C(0, 1), where one of the distractors is presented, are positioned very closely even though they are well separated in the original representation space. This is because the neuron #1 and neuron #2 are particularly responsive to one of the stimuli, but not to the other. Thus, the averaging cancels out these responses to give similarly moderate values. This indicates that a pair of close population averages shown in the last chapter might in fact be very distinctly represented in the original representation space.

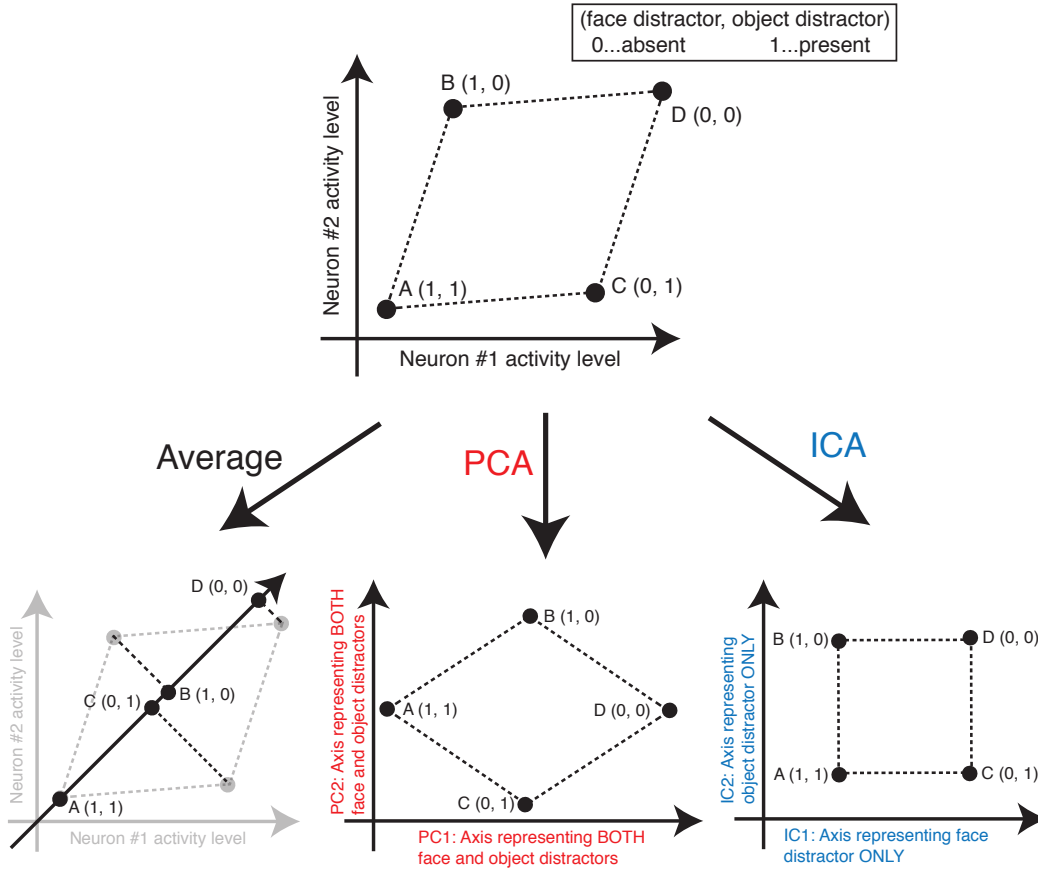


Figure 4.1: Comparison of average, PCA and ICA. The top panel shows hypothetical two face selective neurons responding to a combination of face distractor and/or object distractor, in addition to a preferred face. The four data points A, B, C and D show different stimulus condition with “0” and “1” indicating absence and presence of the (face distractor, object distractor), respectively. For example, B(1, 0) means face distractor is present, but object distractor is absent. The imaginary dotted line is highlighting the shape of the data distribution. Average (across neurons) operation, PCA and ICA are applied to this distribution and presented at the bottom panels. (Left) Result of an average operation. The average operation corresponds to projecting the four points onto the line of unity, and read either x (abscissa) or y (ordinate) value of them (see Appendix H). (Middle) PCA result. PCA finds a new set of axes such that each axis captures as much variance of the distribution as possible, resulting in rotating the original axes. (Right) ICA result. ICA “whiten” the distribution, namely equalize the variances along both dimensions (i.e. both along x and y axes) to find a new set of axes on which marginal distributions have minimum Gaussianity. For simplicity, the distributions are shown with non-zero means along both dimensions (i.e. both along x and y axes) but in the actual PCA and ICA the means of distribution were subtracted to center the distribution on the origin. For the details, see the main text.

4.2 How to capture representation in the high dimensional space spanned by population

However, one issue with treating data without averaging it is visualization. If N cells were recorded with c different stimulus conditions, we need to seek a representation in $N \times c$ matrix, which makes it difficult to grasp the overall response pattern at a glance. One way to solve this problem is to apply dimensional reduction technique, such as Principal Component Analysis (PCA) or Singular Value Decomposition (SVD). For the relationship between PCA and SVD, see Appendix I. These analyses reassign the original axes to the calculated eigenvectors, which capture the largest variance in the data distributed in the representation space. The eigenvectors are calculated from a covariance matrix of the original data, and sorted in descending order of eigenvalues. Each eigenvalue indicates the captured variance in the direction of its corresponding eigenvector. Therefore, eigenvectors associated with small eigenvalues can be safely omitted without large reduction of total explained variance, which means reducing the dimension of the original space, while capturing the principal pattern of activity.

PCA and SVD subtract the mean of the distributed data (center the distribution to the origin) to extract the standard deviation of the distribution [161]. This means that if the data distribution is a multi-dimensional Gaussian (which is defined only by means and standard deviations), PCA can successfully summarize the distribution in the low dimensional space. In this case, marginal distributions along each axis (eigenvector) has a shape of a Gaussian. However, Gaussian marginal distribution can be a problem in terms of decomposing information in the representation space. The reason is as follows: the Central Limit Theorem states that a linear summation of random two samples taken

from two arbitrary distribution will distribute close to Gaussian. In other words, marginal Gaussian distribution indicates that the axis is representing a mixture of information. Fig. 4.1 demonstrates this problem. The bottom middle panel shows the result of PCA applied to the original representation space shown at the top panel. PCA captures the diagonals of original tilted square shaped distribution for its largest standard deviation. The horizontal axis (PC1) has A (0, 0) on the left side close to the origin, B (1, 0) and C (0, 1) on the middle, and D (0, 0) on the right side (which is also similar to the average shown in bottom left panel). Thus, from left to right, both the presence of face distractor stimulus and object distractor stimulus are changing. In other words, this axis represents the mixture of face distractor and object distractor stimuli (the same logic can be applied to the vertical axis). Additionally, because PCA captures the largest variance, the first eigenvector, i.e. PC1 can potentially be very similar to the population average. This can be appreciated by projecting both the population average (Fig. 4.1, left panel) and the PCA result (Fig. 4.1, middle panel) to the horizontal axes. Therefore, PCA is a convenient tool to grasp the shape of the representation, but axes can contain mixed information which often makes data interpretation difficult.

Both averaging and PCA could compromise or mix the information in the original representation space. In order to find axes each of which is representing only one kind of stimulus information, we looked at higher order statistics (i.e. third or higher order) using Independent component analysis (ICA). ICA finds axes whose marginal distribution is non-Gaussian (because a Gaussian distribution indicates mixed information source, as described above) by using higher order statistics such as kurtosis or skewness [162]. Because Gaussian distribution has zero value of higher order statistics, these higher order statistics can be used as a measurement of non-Gaussianity (practically, the absolute

values are used to quantify distance from Gaussian). The bottom right panel of Fig. 4.1 demonstrates the ICA result. In the marginal distribution of the horizontal axis (IC1), A(1, 1), B(1, 0) are on the left, and C(0, 1) and D(0, 0) are on the right. Thus, the horizontal axis is representing the presence of object distractor stimulus, independent of the presence of face distractor stimulus (the same logic is applied to the vertical direction). Another way to see independence in ICA is to determine the data point given a value on one axis. For example, knowing the value of data C or D on the horizontal axes cannot determine one point because data C and D have the common value (i.e. underspecified). However, in the PCA result, knowing the value of data D on horizontal axis defines one point in the data distribution. Thus, PCA axes are not independent, and ICA axes are independent. There are several algorithms for running ICA, including kurtosis-based, entropy-based and mutual information-based, but the underlying principle is similar as above. In the subsequent analyses, we applied Bell-Sejnowski infomax algorithm [163] (MATLAB code based on DTU:toolbox [164]).

4.3 Independent representations of the multiple stimuli in the population activity pattern

According to the thought experiment in the last section, when PCA was applied to the Cat/Num experiment the resulting PCs had mixed information, as shown in the Fig. 4.2. The PC1 coefficients have qualitatively similar pattern as the population average (see Fig. 3.2), also consistent with the thought experiment. PC2 and PC3 seemed to represent mixed information of distractor category, distractor number and mean firing rate.

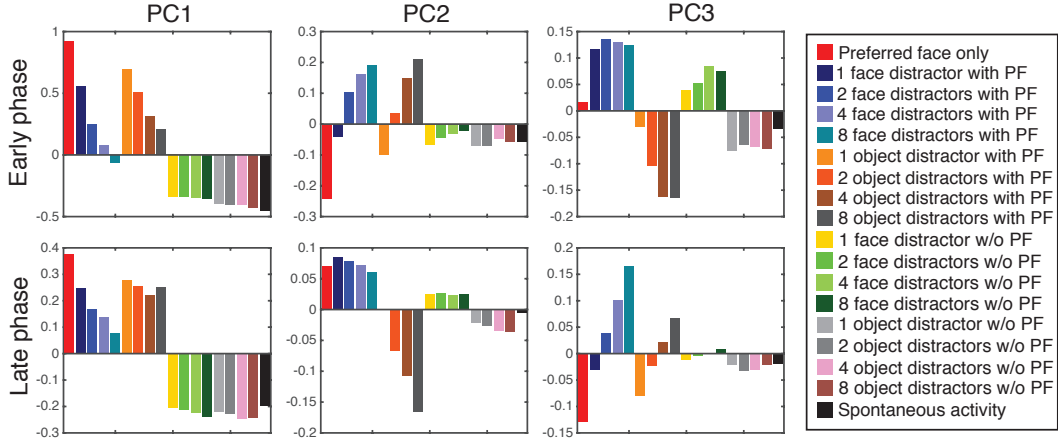


Figure 4.2: PCA applied to the Cat/Num experiment. Unit is arbitrary (normalized firing rate). (Top) Three PCs from the early phase. PC1, PC2 and PC3 captured 83%, 6% and 4% of total variance, respectively. (Bottom) Three PCs from the late phase. PC1, PC2 and PC3 captured 79%, 6% and 6% of total variance, respectively.

We confirmed that PCs have mixed information by running DPCA (Demixed PCA [165]), which could potentially separate stimulus information¹ and ascribe as much as possible to each PC, while maintaining the orthogonality of the axes. When it was applied to the data, DPCA resulted in almost the same result as PCA (thus, DPCA could not completely "demix" the components). The result is shown in Fig. 4.3. The DPC1 is similar to the population average, and all of the DPCs have mixed information (For details, see the caption of Fig. 4.3). One small difference from PCA is that the DPC2 and DPC3 in the late phase are swapped, but since the explained variance by PC2 and PC3 had very similar explained variance in PCA (both 6%, see the caption of Fig. 4.2), we think this result is neither robust nor consistent.

DPCA also showed that each DPC represents a mixture of information: in the early phase for example, 77% of variance in DPC1 was given by either a

¹In DPCA stimulus group was provided by user, thus DPCA is not completely unsupervised algorithm as PCA.

preferred face is either present or absence (variable which is directly related to mean firing rate), but 11% of the variance was explained by the mixture of the existence of a preferred face, distractor number and distractor category. Similarly, while 55 % of DPC2 was explained by difference in distractor number, but 40% of variances are explained by presence of a preferred face, difference in distractor number and category. Also 63% of variance in DPC3 was explained by distractor category difference, while 20% of variance was mixture of presence of a preferred face, difference in distracter number and category.

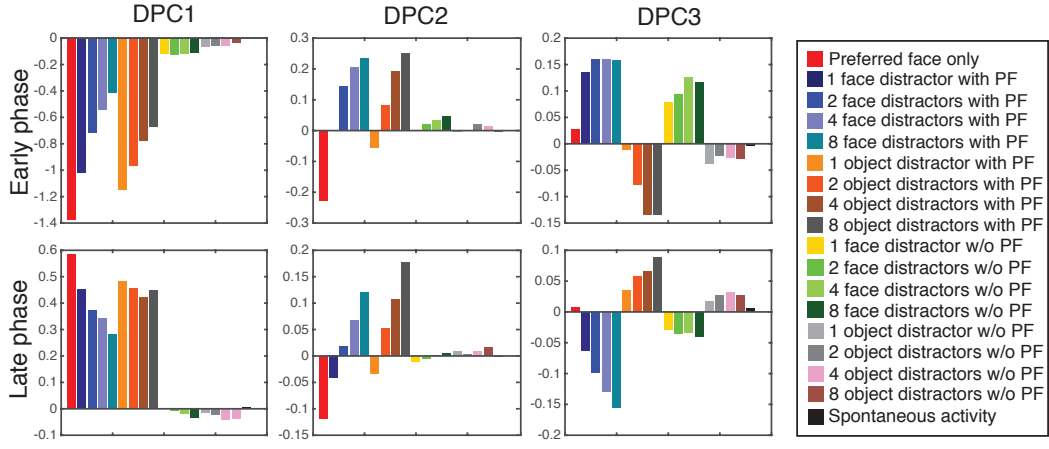


Figure 4.3: DPCA applied to the Cat/Num experiment. Unit is arbitrary (normalized firing rate). (Top) Three DPCs from the early phase. DPC1, DPC2 and DPC3 captured 84%, 6% and 3% of total variance, respectively. (Bottom) Three DPCs from the late phase. DPC1, DPC2 and DPC3 captured 80%, 7% and 5% of total variance, respectively.

To test whether population activity pattern gives axes with independent stimulus information, we applied linear ICA to the population activity matrix. Because ICA does not reduce the dimension of matrix², SVD was used first to reduce the original N (number of recorded neuron) dimension to k dimensions

²To be precise, PCA also does not reduce the dimension per se, but variance explained by given PCs are in descending order such that a user can discard PCs with small explained variance - resulting in reducing dimension - without compromising total explained variance. ICA on the other hand does not provide ICs with decaying explained variance.

(k is an arbitrary number, determined by user based on number of information needed to be represented). Then, ICA was applied to find s independent components across which minimal mutual information was contained. The analysis streamline is illustrated in Fig. 4.4.

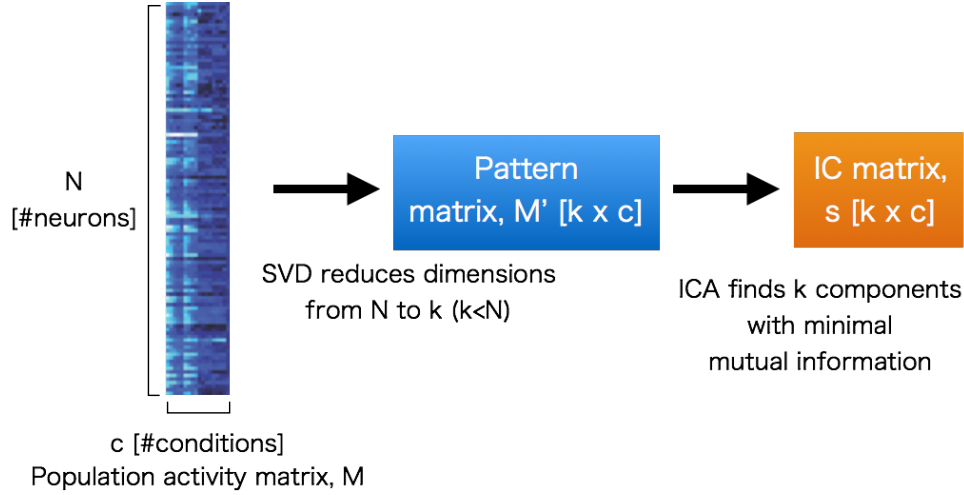


Figure 4.4: The concept of dimension reduction followed by ICA. The original response matrix was $N \times c$ two-dimensional matrix, where N is number of the recorded neurons and c is the number of stimulus conditions. Matrix elements were time average of either early or late phase response. SVD reduced the first dimension from N to either 5 (Cat/Num experiment) or 3 (Prox/Ecc experiment). From this dimension-reduced space, ICA found independent components based on Bell-Sejnowski Infomax algorithm.

We found independent components represent stimulus parameter we used in the Cat/Num experiment (Fig. 4.5). IC1 represented the distractor number. Note that sign and amplitude (i.e. bar graph direction and absolute height) is irrelevant in ICs³. Thus, only relative heights of the bars are important. The four blue (face distractors and preferred face conditions) bars and yellow bars (object distractors and preferred face conditions) look identical, indicating that IC1 was counting the distractor independent of distractor category. However,

³In ICA, the original data is reproduced as a linear combination of an IC matrix and a mixing matrix. Because both of the IC matrix and mixing matrix, multiplying an IC with a scalar value can be cancelled out by dividing a mixing matrix by the scalar value (and *vice versa*). Therefore, the magnitude and signs of the ICs are underspecified.

IC2 and IC3 depended on distractor category. IC2 had peaks only at the preferred face and object distractors condition (the 6 - 9th bars from the left, yellow-brown color group). Similarly, IC3 had peaks at preferred face and face distractors conditions (the 2 - 5th bars from the left, blue - green bar group). Note that the ICs are ordered according to their explanatory power, similar to the explained variance in PCA (responses to the preferred face and face distractors were smaller than responses to the preferred face and object distractors, therefore the latter represented IC2). IC4 and IC5 represented face and object distractors without the preferred face, respectively. As described in the last chapter, the responses without the preferred face were very weak at the early phase, and almost same as baseline firing rate in the late phase. Moreover, there were almost no magnitude differences between face and object distractors. Nevertheless, IC4 and IC5 discriminated face and object distractors, both at the early and the late phase. This finding was similar when we varied the distractor proximity. Regardless of the phase, the stimulus information was represented in the ICs. IC1 represented the preferred face (without distractors), IC2 represented the distractor distance at the preferred face with distractor condition, and IC3 represented the distractor distance at the distractor only control condition. However, when we varied the eccentricity of stimulus position, we found slightly different representation at the RF center and RF periphery. At the RF center (Fig. 4.5B), preferred face (without distractor, the red bar), preferred face with distractors (the blue bars, 2 - 4th from the left), distractors without preferred face (the green bars, 5 - 7th from the left) were separated in IC1, IC2 and IC3, respectively. However, at the RF periphery, preferred face with distractors and preferred face without distractors are grouped together (Fig. 4.5C)⁴. This may indicate that at the

⁴For example, IC2 in Fig. 4.5 has similar peaks at “preferred face with medium distance distractors”, “preferred face with small distance distractors”, “medium distance distractors

RF periphery the stimulus information could be compromised. Thus, although minor differences were present, ICs in the early and late phase revealed almost the same information contents, despite the large magnitude difference in the mean firing rate of the early and the late phase.

4.4 Quantifying discriminability assessing temporal transition of representation

ICA revealed the representation of stimulus information in the pattern of population activity. This showed that the neural population was able to discriminate each stimulus condition, instead of its representation being compromised by the distractors. To quantify the detectability, we ran a supervised algorithm on the same data (using Neural Decoding Toolbox [166]) to study whether distractor category, distractor number and distractor proximity could be decoded from the response. The algorithm used was machine learning based on a population vector calculated in the training phase (either the early or the late phase response were used), and tested within the same phase (for detail, see Appendix A). As described in Chapter 3, the early phase and the late phase were defined such that both phase have approximately equal number of spikes. This design allowed us to quantify information content, or discriminability per unit spike in each phase (because each phase is defined such that they have approximately the same number of spikes). Fig. 4.6A shows the decoder performance. The discriminabilities were above chance level in both the early and in the late phase, and the difference in performances were not significant in the early and the late phase, although response magnitudes were significantly

without preferred face” and “small distance distractors without preferred face”, may indicate that the IC2 is mixing stimulus conditions where distractors have certain distance, regardless of presence of the preferred face.

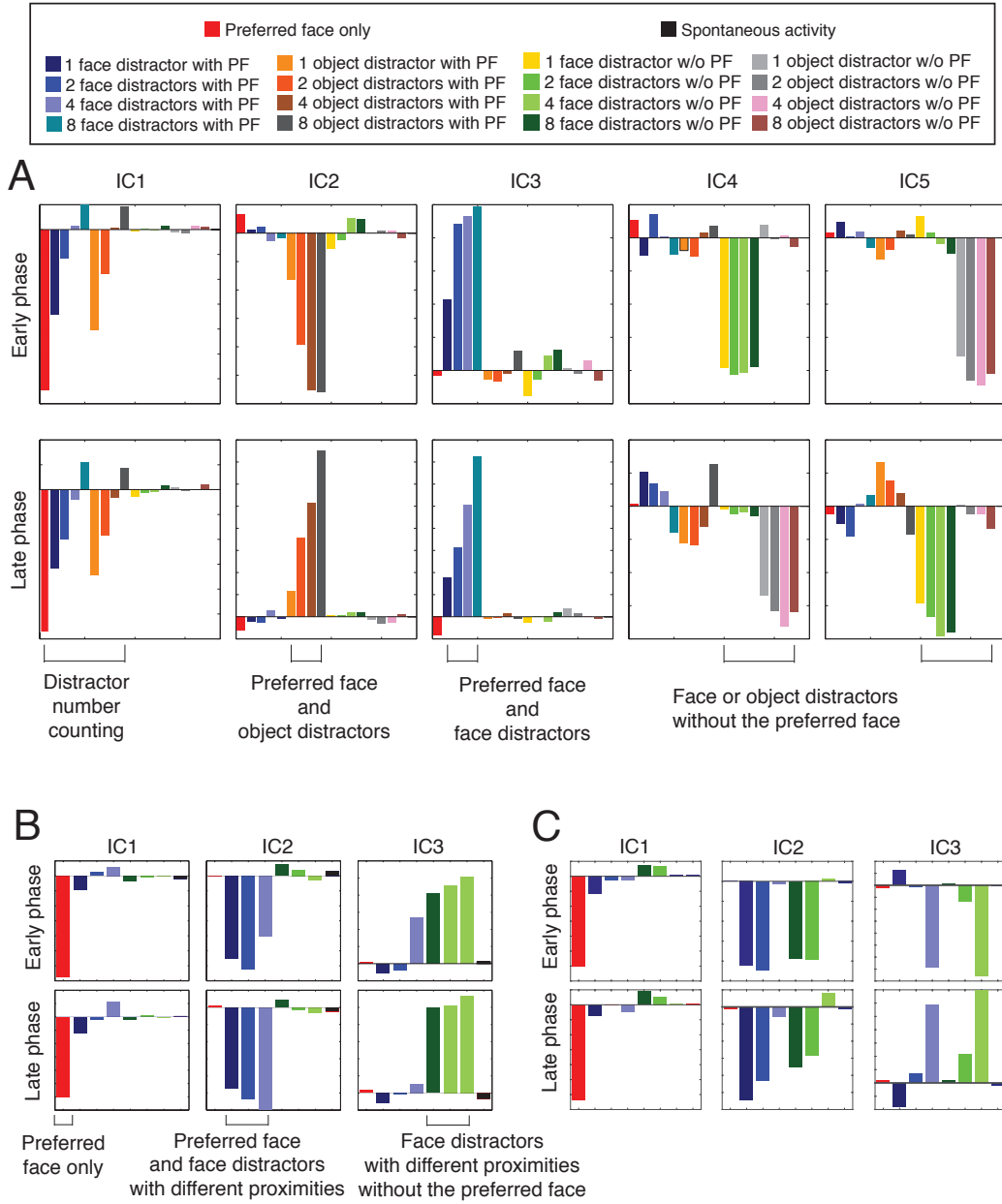


Figure 4.5: High-dimensional representation preserves stimulus discriminability. Dimensions originally spanned by neural population are reduced to 5 or 3 using SVD followed by ICA. Color code and abbreviation follows Fig. 3.2 and 3.6. (A) Cat/Num experiment. Each bar corresponds to IC value of the 18 conditions used in Fig. 3.2. (B) Prox/Ecc experiment at RF center. Each bar corresponds to IC value of the 7 conditions used in Fig. 3.6. (C) Same as (B), but the stimuli were presented at the periphery of RF.

different as we saw in Chapter 3. This led us to another speculation that the neural coding scheme might have changed from the early to the late phase.

As a first attempt to compare the coding scheme in the early phase and the late phase, we ran the temporal cross-training analysis (also in Neural Decoding Toolbox, Fig. 4.6B). In this analysis, instead of training and testing a machine in the same time frame, a machine is tested in one particular window, while trained in all time windows. As seen in the Fig. 4.6B, we found a two-block pattern: a small in the block upper left corner (100~200 ms after the stimulus presentation) corresponding to the early phase, and a larger bottom right block (250~400 ms after the stimulus presentation) corresponding to the late phase. This two block pattern indicated that within the early and late phase the coding scheme was similar (in other words, one can train and test the machine within different time windows to obtain a high decoding performance), but the coding scheme was different from the early to the late phase (otherwise we should see one large block pattern). Then, how did the encoding scheme change? We explore this question further in the next chapter using a computer model.

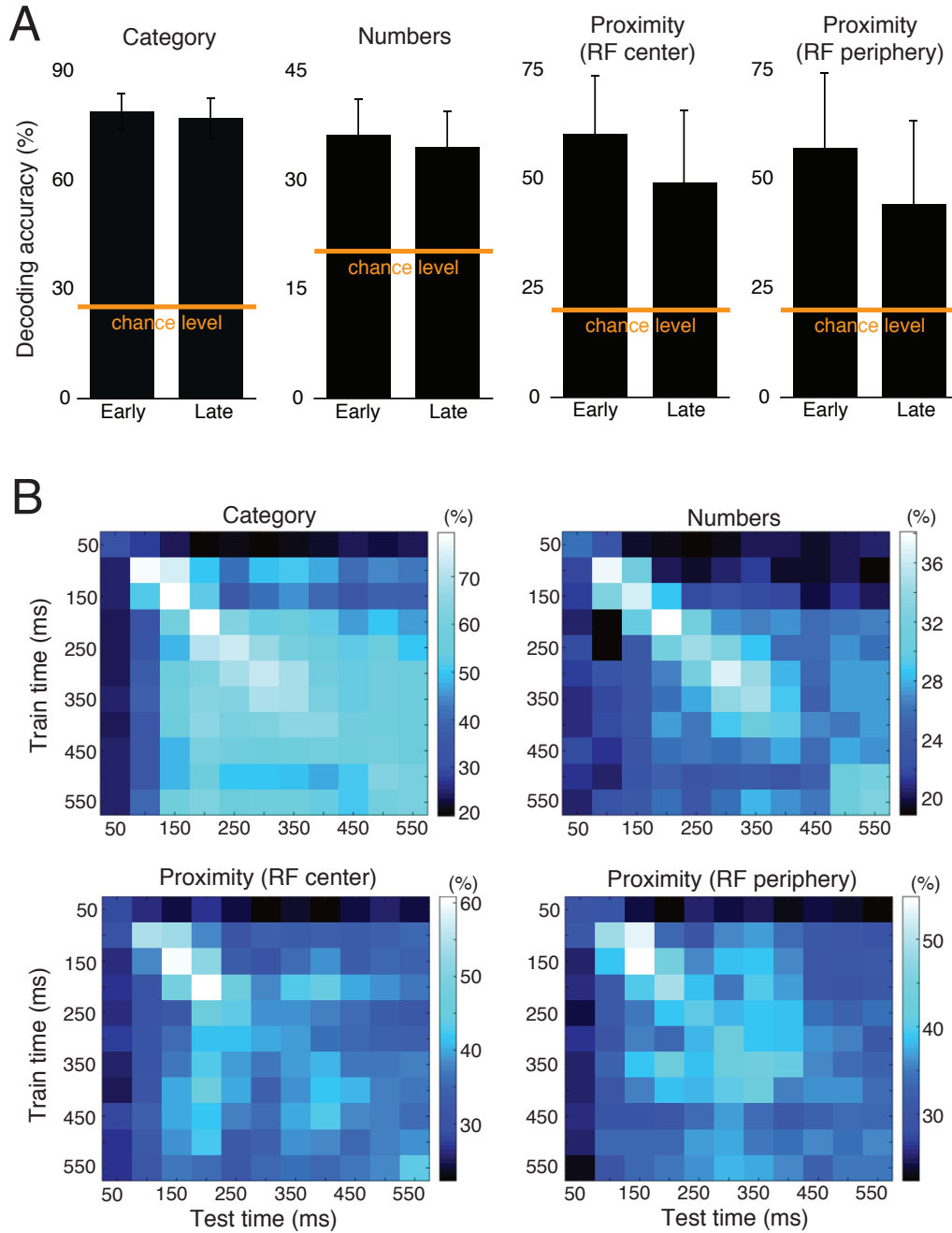


Figure 4.6: Decoding analysis on population activity. The machine learning algorithm based on correlation between population vectors was used to quantify decoding accuracy (Neural Decoding Toolbox [166]). (A) From left to right, decoding results of distractor category, distractor numbers, proximity at RF center and proximity at RF periphery. For category decoding, stimuli with different distractor numbers were grouped together as long as distractor category was same (i.e. face or object). Similarly, for number decoding, stimuli with different distractor categories were grouped together. Chance levels are shown with the orange horizontal line. Error bars are standard deviations over 50 resample runs. (B) Temporal-cross-training analysis. Classifier (machine) was trained at a time window specified by the ordinate, and tested at a time on the abscissa.

5

Modeling results

5.1 Divisive Normalization framework to explain discriminability

With unsupervised (ICA) and supervised (machine learning) learning algorithm, we showed that stimulus information - distractor number, distractor category and distractor proximity - were preserved in the population activity. Moreover, the machine learning study indicated that the information encoding scheme was changed from the early to the late phase. The next question is, HOW did the encoding scheme change across time? To study the stimulus encoding further, we developed a computer model based on Divisive Normalization.

Why is Divisive Normalization suitable for our data? By fMRI data, we know that temporal lobe contains both face selective cortices and non-face selective cortices. This indicates that presenting face or non-face object produces large population activity in each of the selective cortices. In terms of Divisive Normalization, large population activity induces large normalization.

Moreover, as described above, Divisive Normalization can bias the weight in proportion to the population activity. This indicates that, presenting a preferred face and a non-face distractor induces a biased weight toward the face within the face patch population (because neural population is selective to face), resulting in slight response reduction (i.e. closer to winner-take-all). On the other hand, presenting a preferred face and a face distractor induces a larger normalization power with almost equal weights (because both of the two presented images are faces, which are in the selective category of the face patch), resulting in larger response reduction. In fact, this is exactly what we observed in our data: face distractors suppressed responses more than objects. Face distractors themselves do not elicit a response to the neuron from which we are recording, but they do elicit a large response in the surrounding population which normalizes the response of the neuron that we are observing. In contrast, object distractors do not elicit large response to the neuron, and also do not elicit large response to the neighboring population either, resulting in small normalization from object distractors.

5.2 Divisive normalization model explained the early phase of Category/Number variation experiment without any free parameter

Based on the Divisive Normalization framework, we developed several models. Firstly, I introduce the pure prediction model. The name “pure prediction” came from the fact that this model does not contain any free parameter. Thus, this model is directly testing our working hypothesis without fitting the model behavior to the observed data. The model equation is as follows:

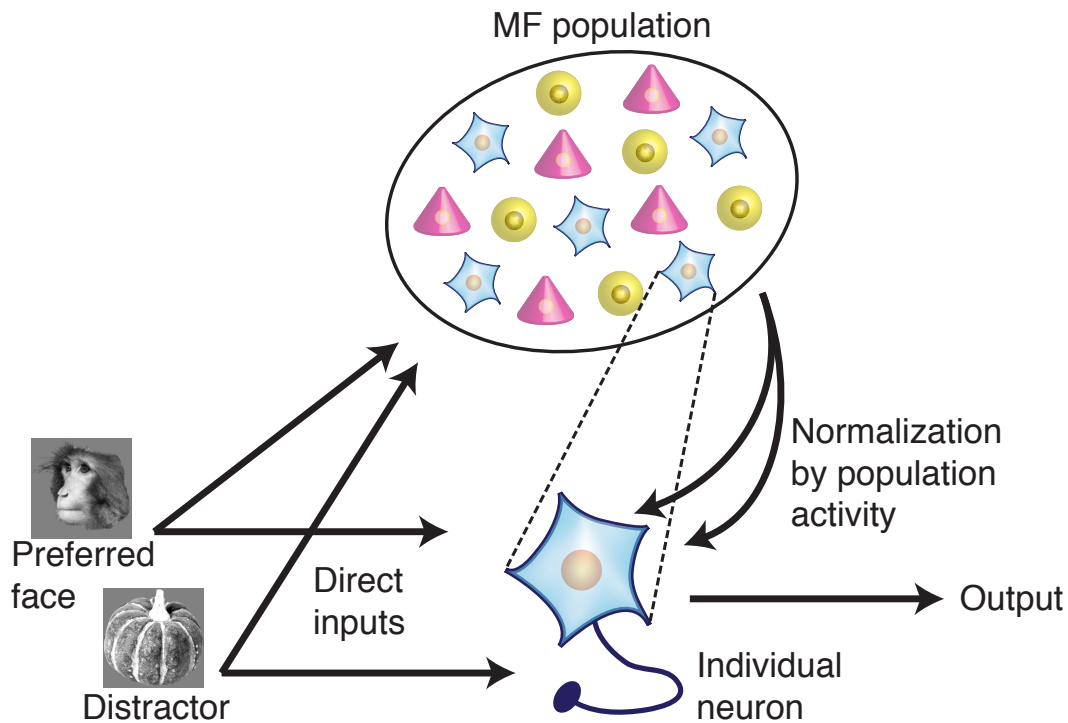


Figure 5.1: The concept of Divisive Normalization. Divisive Normalization states that a neuron computes a ratio between the direct inputs and population activity of neighboring neurons. Even though we selected a non-preferred face for the recorded neuron, there would be a certain fraction of neurons activated by the face in the neighboring population (a non-preferred face for one neuron might be a preferred face for the other). This indicates a large normalization induced by non-preferred faces. However, for non-preferred object this may not be the case, because face patch contains more than 90 % face selective neurons, resulting in a small normalization.

$$S_i^{\text{pref+dist}} = \frac{\sum_{p=1}^l S_i^{\text{pref}}(x_p) P_i^{\text{pref}}(x_p) + \sum_{q=1}^m S_i^{\text{dist}}(y_q) P_i^{\text{dist}}(y_q)}{\sum_{p=1}^l P_i^{\text{pref}}(x_p) + \sum_{q=1}^m P_i^{\text{dist}}(y_q)} \quad (5.1)$$

S_i^{pref} and S_i^{dist} in the numerator are direct inputs S from the preferred face and the distractor, respectively. The actual value used for these direct inputs were calculated based on measured RF value. x and y are specifying the presentation location on the screen. These direct inputs were weighted and divided by population activity, P . These population activities were calculated by average RF of the stimulus used (for details of calculation, see Appendix A.). When there are multiple preferred faces or distractors, the number of terms in the denominator and numerator are increased. We applied this model to all of the recorded neurons and averaged across the neurons. As shown in Fig. 5.2, the model and data show a fairly close correspondence within the error bar range (SEM). Two-way ANOVA did not show significance between the data and prediction for the 18 stimulus conditions shown in Fig. 5.2 and Table 5.1 ($F(1, 3960) = 0.67, p = 0.41 > 0.05$).

This pure prediction model worked particularly well at the early phase of the distractor numbers / category variation experiment, indicating that Divisive Normalization could predict the neural computation scheme in this condition (i.e. the early phase of the distractor numbers / category variation experiment).

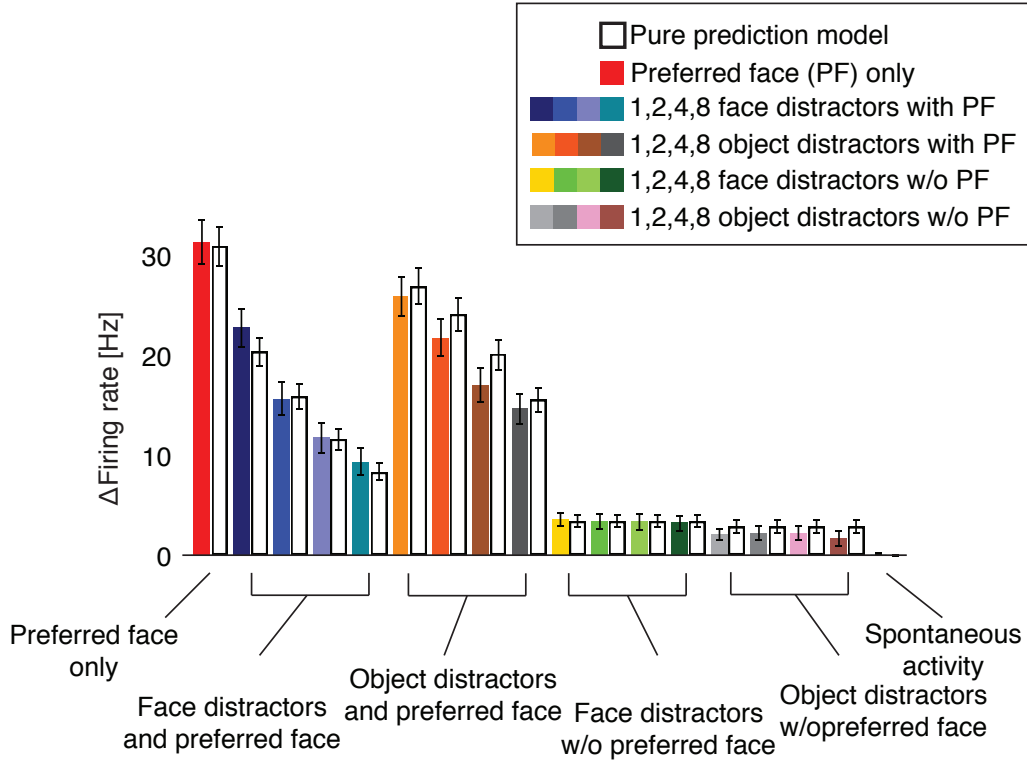


Figure 5.2: Pure prediction model showed close correspondence to the independently measured data. Mean firing rates and model outputs were calculated at the early phase and averaged across recorded neurons (colored and open bar graphs, respectively), using data from Cat/Num experiment. The color code follows Fig. 3.2. Error bars show standard error of the mean.

Table 5.1: Data and the model average response and SEM

	Data	Model
PF.	31.41 ± 2.24	31.00 ± 2.04
PF+1FD.	22.82 ± 1.99	20.41 ± 1.52
PF+2FD.	15.69 ± 1.70	15.92 ± 1.28
PF+4FD.	11.83 ± 1.53	11.68 ± 1.05
PF+8FD.	9.37 ± 1.39	8.37 ± 0.88
PF+1OD.	26.00 ± 2.05	27.00 ± 1.85
PF+2OD.	21.86 ± 1.90	24.12 ± 1.72
PF+4OD.	17.09 ± 1.67	20.17 ± 1.53
PF+8OD.	14.73 ± 1.56	15.62 ± 1.30
1FD.	3.63 ± 0.72	3.43 ± 0.69
2FD.	3.47 ± 0.79	3.43 ± 0.69
4FD.	3.46 ± 0.85	3.43 ± 0.69
8FD.	3.26 ± 0.83	3.43 ± 0.69
1OD.	2.15 ± 0.62	2.93 ± 0.69
2OD.	2.27 ± 0.74	2.93 ± 0.69
4OD.	2.24 ± 0.77	2.93 ± 0.69
8OD.	1.73 ± 0.78	2.93 ± 0.69
S.A.	0.14 ± 0.15	0 ± 0

5.3 Limitation of the pure prediction model and introducing an exponent as a free parameter

However, for Prox/Ecc experiment, and in the late phase, the pure prediction model gave an error (Fig. 5.3). In the late phase, some neurons showed responses different from the early phase: for example, response became robust against distractors and response reduction was decreased. In Prox/Ecc experiment, although Divisive normalization model could capture the qualitative reduction of firing rate as distractor distances were decreased, but the model overestimated the response suppression to have a much smaller firing rate even when the distractor distance was longest (Fig. 5.3, open bars) In or-

der to understand the computational principle in the rest of the conditions, we introduced the post-diction model. The model equation is same as the widely used, common Divisive normalization model:

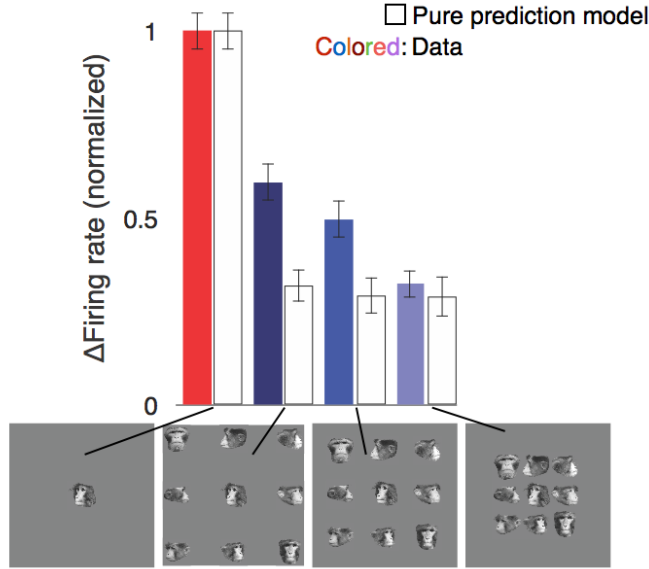


Figure 5.3: Pure prediction model showed a larger error for the Prox/Ecc experiment. Mean firing rates and model outputs were calculated at the early phase and averaged across recorded neurons, using data from Prox/Ecc experiment. The color code follows Fig. 3.6.

In this post-diction model (Equation (5.2)), a free parameter n is used. This free parameter is an exponent on the population activity allows one to change the bias of the weight either toward one direct input or another. Note that in this configuration the sum of coefficients on the direct inputs are always equal to 1, allows easier quantification of the computational bias.

$$S_i^{\text{pref+dist}} = \frac{\sum_{p=1}^l S_i^{\text{pref}}(x_p) \{P_i^{\text{pref}}(x_p)\}^n + \sum_{q=1}^m S_i^{\text{dist}}(y_q) \{P_i^{\text{dist}}(y_q)\}^n}{\sum_{p=1}^l \{P_i^{\text{pref}}(x_p)\}^n + \sum_{q=1}^m \{P_i^{\text{dist}}(y_q)\}^n} \quad (5.2)$$

5.4 Divisive Normalization model revealed the change of computational principle from the early to the late phase

With the post-diction model, all the remaining conditions were explained (Fig. 5.4). Not only the mean firing rate, but also the results of population readout analyses by ICA were reproduced. With the post-diction model, we are now ready to address the question: how did the stimulus coding scheme change across time?

For simplicity, let us describe the post-diction model in case of the two stimuli condition, namely when only one preferred face and one distractor (either a face or an object) is presented on the screen. The post-diction model can be written as a linear summation of two direct inputs:

$$S_i^{\text{pref+dist}} = \frac{\{P_i^{\text{pref}}(x_p)\}^n}{\{P_i^{\text{pref}}(x_p)\}^n + \{P_i^{\text{dist}}(y_q)\}^n} S_i^{\text{pref}}(x_p) + \frac{\{P_i^{\text{dist}}(y_q)\}^n}{\{P_i^{\text{pref}}(x_p)\}^n + \{P_i^{\text{dist}}(y_q)\}^n} S_i^{\text{dist}}(y_q) \quad (5.3)$$

As described above, two coefficients are normalized such that the sum of

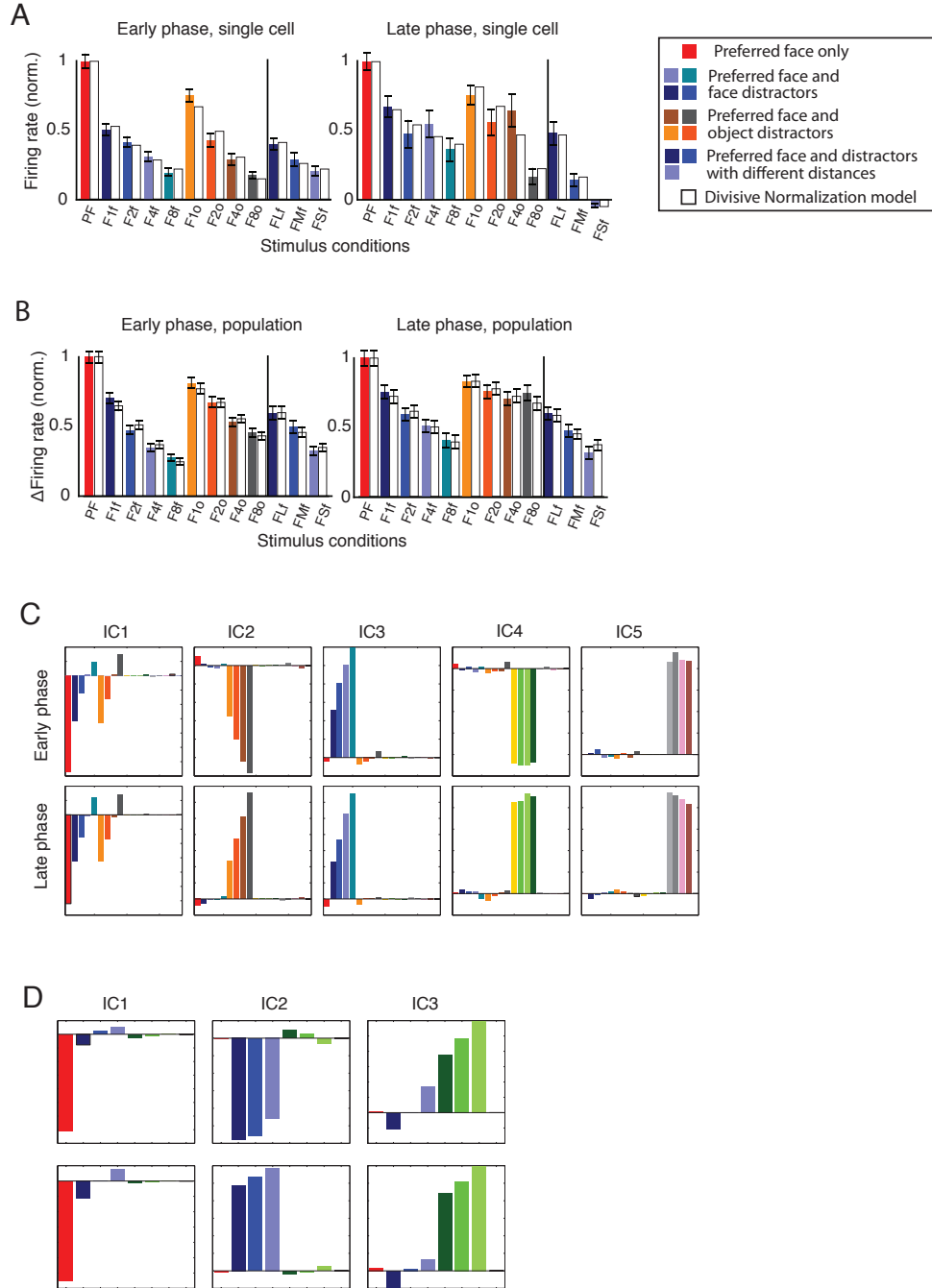


Figure 5.4: Post-diction model explains the observed representation. Model outputs of single cell response, population average and high-dimensional space representation reproduced the experimental observation. Color code and abbreviation follows the Fig. 5.2. (A,B) The post-diction model (white bars) compared to both Cat/Num and Prox/Ecc experiment, at the early and the late phase. (A) shows a single cell example, and (B) shows the population average across recorded neurons. (C,D) Post-diction model of Cat/Num and Prox/Ecc experiment, respectively, analyzed with ICA.

them are equal to 1. This means that from the nature of one of the coefficients, we can infer the way in which the neuron combines its inputs and performs the computation. For example, if the coefficient of the first term (preferred face) is equal to 1, the neuron is ignoring the distractor because the coefficient of the second term is 0. If both coefficients are 0.5 however, the neuron is taking the average of the two inputs. Since the two coefficients are complementary, we focused on the first coefficient and named it distractor robustness index.

$$RobustnessIndex = \frac{\{P_i^{\text{pref}}(x_p)\}^n}{\{P_i^{\text{pref}}(x_p)\}^n + \{P_i^{\text{dist}}(y_q)\}^n} \quad (5.4)$$

We calculated the distractor robustness index across all the recorded neurons and created a distribution histogram. Fig. 5.5A shows that at the early phase of the response to the preferred face and a face distractor, most of the neurons had a value of 0.5, indicating average response. However, at the late phase, a significant fraction of the neurons changed their weight toward 1, indicating they started ignoring the distractors. Moreover, this effect was even more prominent in the preferred face and object distractors condition (Fig. 5.5B). At the early phase, the distribution was almost flat, but at the late phase most of the neuron had a value of 1, indicating robustness against the distractors.

In the Prox/Ecc experiment, this tendency also held true. Fig. 5.6 shows that more neurons shifted their index value toward the late phase, and the effect is stronger when the distractor distance was longer. This is reminiscent of the result in Chapter 3, where neurons were not robust against the short distance distractors, resulting in reduced neural response.

By analyzing the distractor robustness index, we found the global shift from averaging regime toward robust regime in both Cat/Num and Prox/Ecc

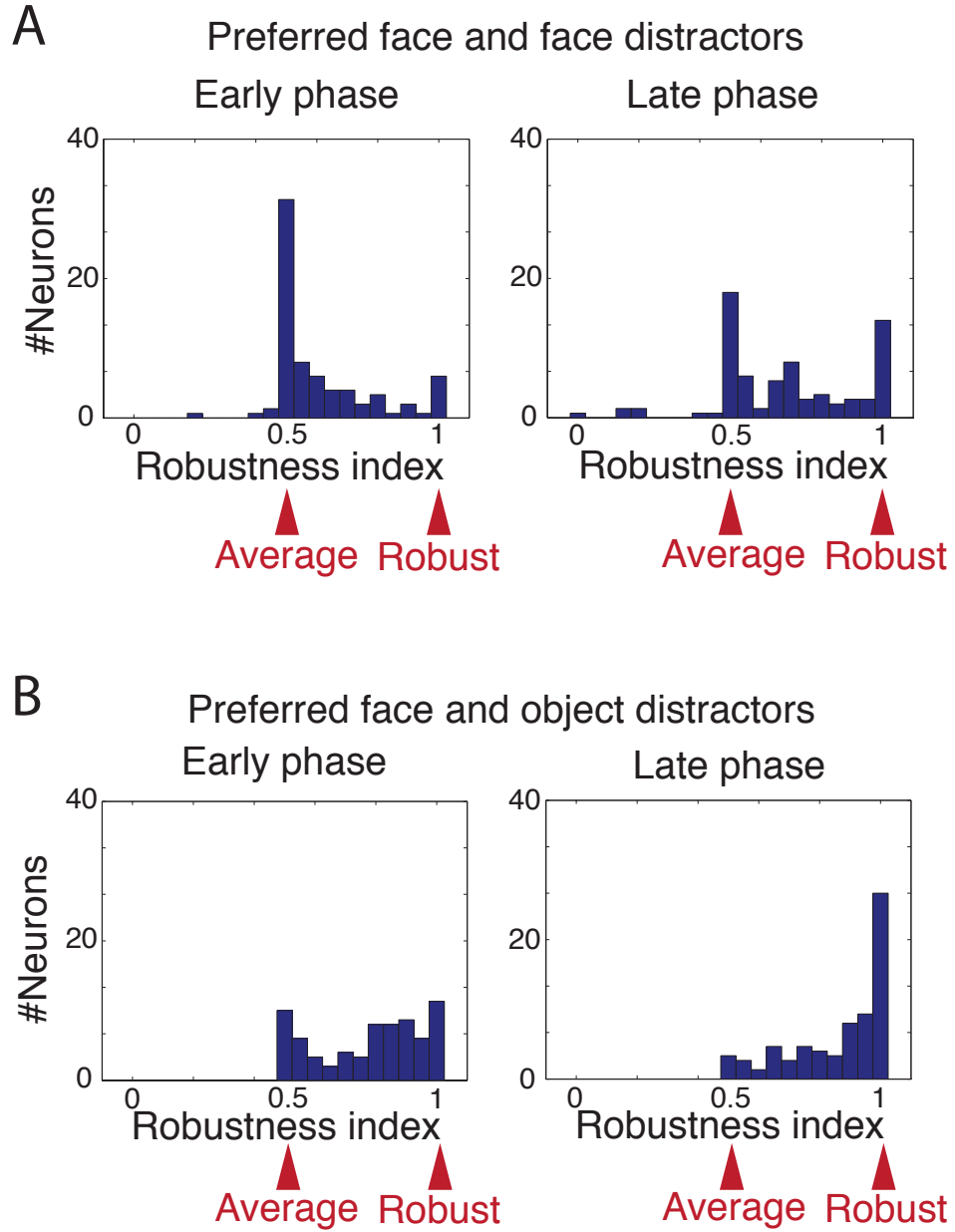


Figure 5.5: Robustness index calculated for the Cat/Num experiment. Each histogram shows the number of neurons in a given range of robustness index. Red arrowheads show that the pointed value of robustness index indicates average or robust computation. (A) Preferred face and face distractor conditions. Left: early phase, right: late phase. (B) Preferred face and object distractor conditions. Left: early phase, right: late phase.

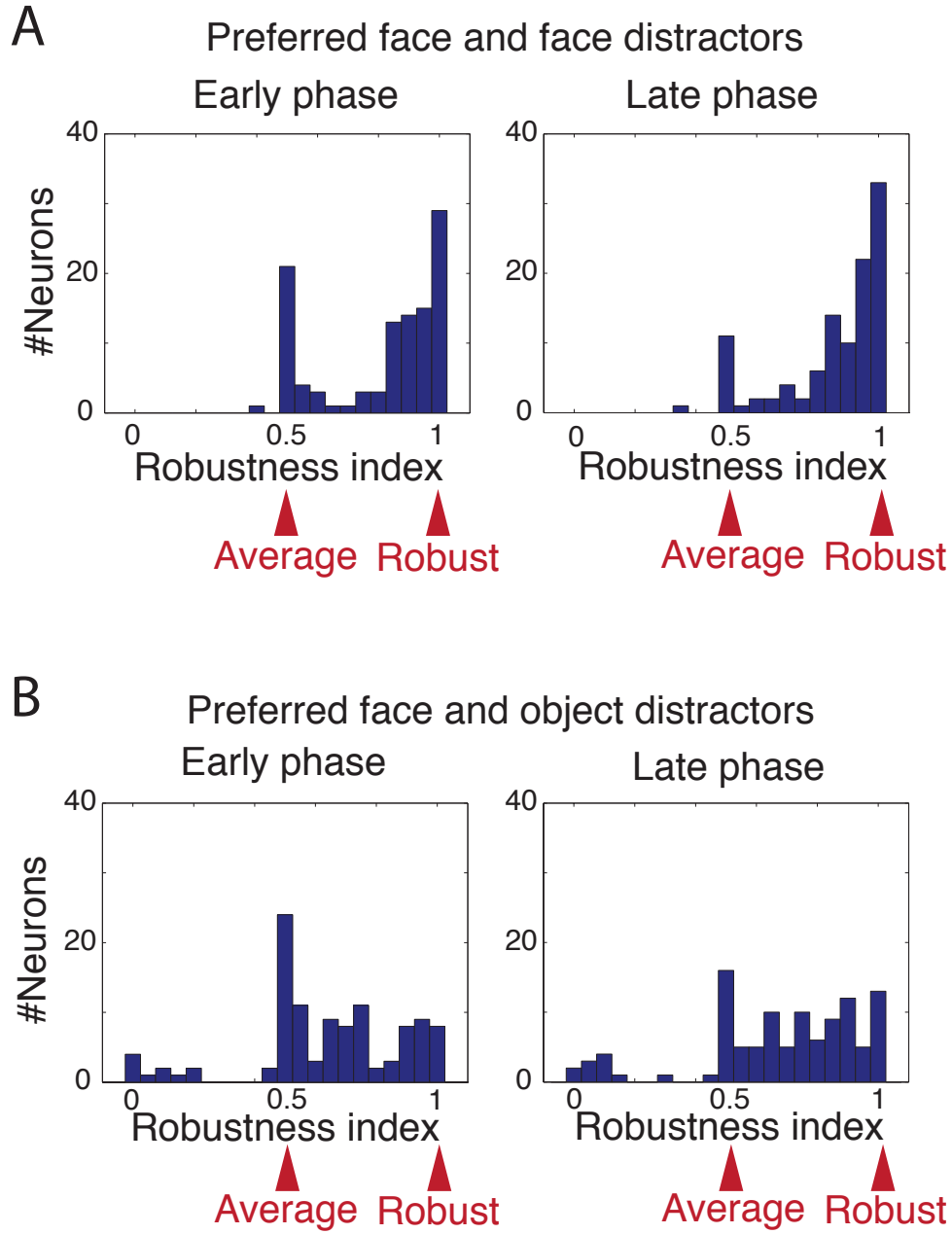


Figure 5.6: Robustness index calculated for the Prox/Ecc experiment. Each histogram shows the number of neurons in a given range of robustness index. Data from RF center presentation is shown. (Top row) Large distractor distance condition. Left: early phase, right: late phase. (Bottom row) Small distractor distance condition. Left: early phase, right: late phase.

experiments. However, as seen in Fig. 5.5 and 5.6, there were always a certain numbers of neurons between the averaging and robust regime. As shown in Chapter 4, this heterogeneous neural population could create independent representation axes in the high dimensional representation space. In fact, we found that the heterogeneity in activity pattern was an important component to have representation axes discriminating the stimuli, as explained below. Fig. 5.7 shows a simulation result of two neural population of 100 neurons, one has homogeneous and the other has heterogeneous response (Fig. 5.7A,B). Here, simplified version of pure prediction models are used to create simulated responses for face distractor conditions (i.e. preferred face and 1, 2, 4 or 8 face distractors, Equation 5.5) and object distractor conditions (i.e. preferred face and 1, 2, 4 or 8 object distractors, Equation 5.6):

$$S_i^{\text{pref+dist,face}} = \frac{S_i^{\text{pref}} P_i^{\text{pref}} + n S_i^{\text{dist,face}} P_i^{\text{dist,face}}}{P_i^{\text{pref}} + n P_i^{\text{dist,face}}} \quad (5.5)$$

$$S_i^{\text{pref+dist,object}} = \frac{S_i^{\text{pref}} P_i^{\text{pref}} + n S_i^{\text{dist,object}} P_i^{\text{dist,object}}}{P_i^{\text{pref}} + n P_i^{\text{dist,object}}} \quad (5.6)$$

where $n = 0, 1, 2, 4, 8$ is distractor number. The single neuron responses, S_i^{pref} , $S_i^{\text{dist,face}}$ and $S_i^{\text{dist,object}}$ and the population response to a preferred face, P_i^{pref} are fixed while the population responses to distractors, $P_i^{\text{dist,face}}$ and $P_i^{\text{dist,object}}$ are also fixed for simulating homogeneous population, and randomly drawn from Gaussian distributions for simulating heterogeneous population. Robustness index calculated from these two population confirmed homogene-

ity (Fig. 5.7B) and heterogeneity (Fig. 5.7F), respectively. Stimulus conditions from Cat/Num experiment were used: preferred face only, preferred face with 1,2,4 or 8 distractors, preferred face with 1,2,4 or 8 object distractors. Population average across neurons showed very similar pattern as shown in the right hand side of Fig. 5.7A and E. However, when ICA was applied to these two simulated population, the result was quite different. The homogeneous population gave almost identical ICs (due to the redundancy in activity pattern), shown in Fig. 5.7C. Moreover, representation of distractor category and numbers are mixed together and ICs could not provide more representation than what average response provides. When ICA was applied to heterogeneous neural population, result showed similar pattern as the recorded data (Fig. 4.5, 5.7G). IC1, IC2 and IC3 were representing distractor number, object distractor and face distractor. PCA results also showed different results in homogeneous and heterogeneous population (Fig. 5.7D,H). The first PC from both population has a similar pattern as population average, But the PC2 from heterogeneous population resembles PC3 (early phase) and PC2 (late phase) from the data (Fig. 4.2), also highlighting the advantage of heterogeneity.

Heterogeneous response pattern in a neural population allowed discriminating stimulus groups in population readout. However, only with the heterogeneity, the groups cannot be discriminated in mean firing rate. Fig. 5.8 gives an example. Fig. 5.8A and E show simulated heterogeneous neural population of 100 neurons. The difference between Fig. 5.8A and E is that the population response (P_i^{dist}) to face distractors is greater than response to object distractors in E, while in A population response shows an equal magnitude of response to face and object distractors. The direct inputs to single neurons are same across neurons and across distractors (i.e. face distractor and object distractor gives the same direct input). As Fig. 5.8B and F shows, population readout

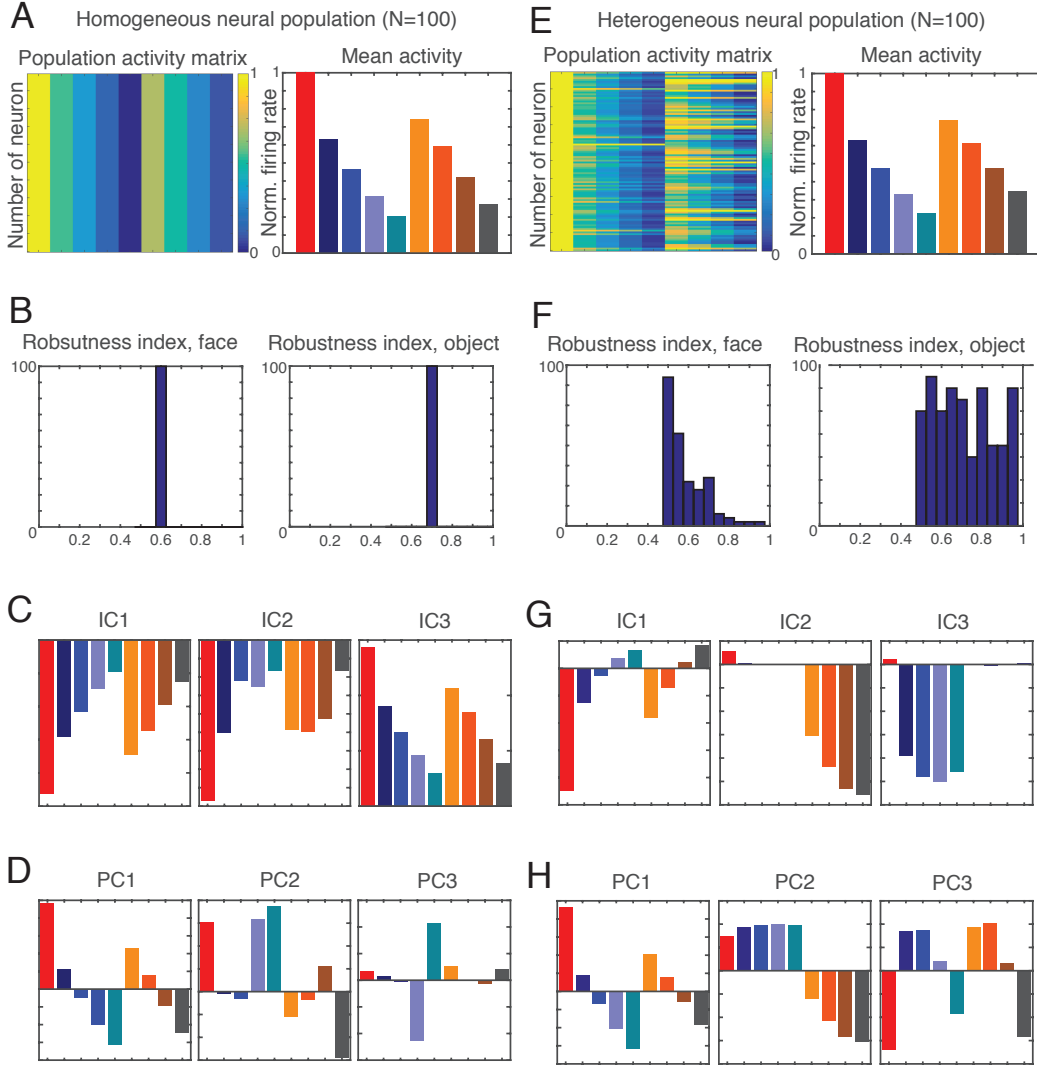


Figure 5.7: Simulation of homogeneous and heterogeneous neural population. Population size is $N = 100$ and stimulus conditions are from Cat/Num experiment. (A, B, C, D) Homogeneous neural population. (E, F, G, H) Heterogeneous neural population. (A) Left, Population activity matrix of the homogeneous population. Right, average neural activity across neurons. (B) distractor Robustness Index distribution across homogeneous population. Left, preferred face and face distractor conditions, Right, preferred face and object distractor conditions. (C) ICA applied to the homogeneous population activity matrix in (A). (D) PCA applied to the homogeneous population activity matrix in (A). (E) Left, population activity matrix of the heterogeneous population. Right, average neural activity across neurons. (F) distractor Robustness Index distribution across heterogeneous population. Left, preferred face and face distractor conditions, Right, preferred face and object distractor conditions. (G) ICA applied to the heterogeneous population activity matrix in (E). (H) PCA applied to the heterogeneous population activity matrix in (E). Color code follows the Fig. 5.2.

of these two simulated population resulted in quite similar IC patterns representing distractor number (IC1), preferred face and object distractors (IC2) and preferred face and face distractors (IC3). However, the mean firing rate of population cannot discriminate face distractor groups (Shown in Fig. 5.8A, right panel. 2 - 5th bar graphs from the left) and object distractor groups (Shown in Fig. 5.8A, right panel. 6 - 9th bar graphs) in Fig. 5.8A. On the other hand, face distractors reduced firing rates more than object distractors in Fig. 5.8E, similarly as the recorded data. Also, the left panel of Fig. 5.8A and E shows that neurons are discriminating the face and object distractor conditions in single neuron level. Thus, the different population response to the different category of distractors can allow neurons to discriminate stimulus conditions, even when the direct inputs from face and object distractors are the same. Face neurons are segregated into modular structure (i.e. face patch), thus population of the face patch shows larger activity to face stimulus more than the object stimulus. As shown in Fig. 5.8E, this difference in population activity resulted in discriminating distractor category in mean firing rate. In other words, modularity helps discriminating different stimulus category both in single neuron level, and population average firing rate level.

We found that heterogeneity in the population activity could represent stimulus information separable by ICA. Then, what type of heterogeneity can help representation and discrimination of stimuli? We discuss relationship between the heterogeneity in MF neurons and Divisive Normalization in Chapter 7.

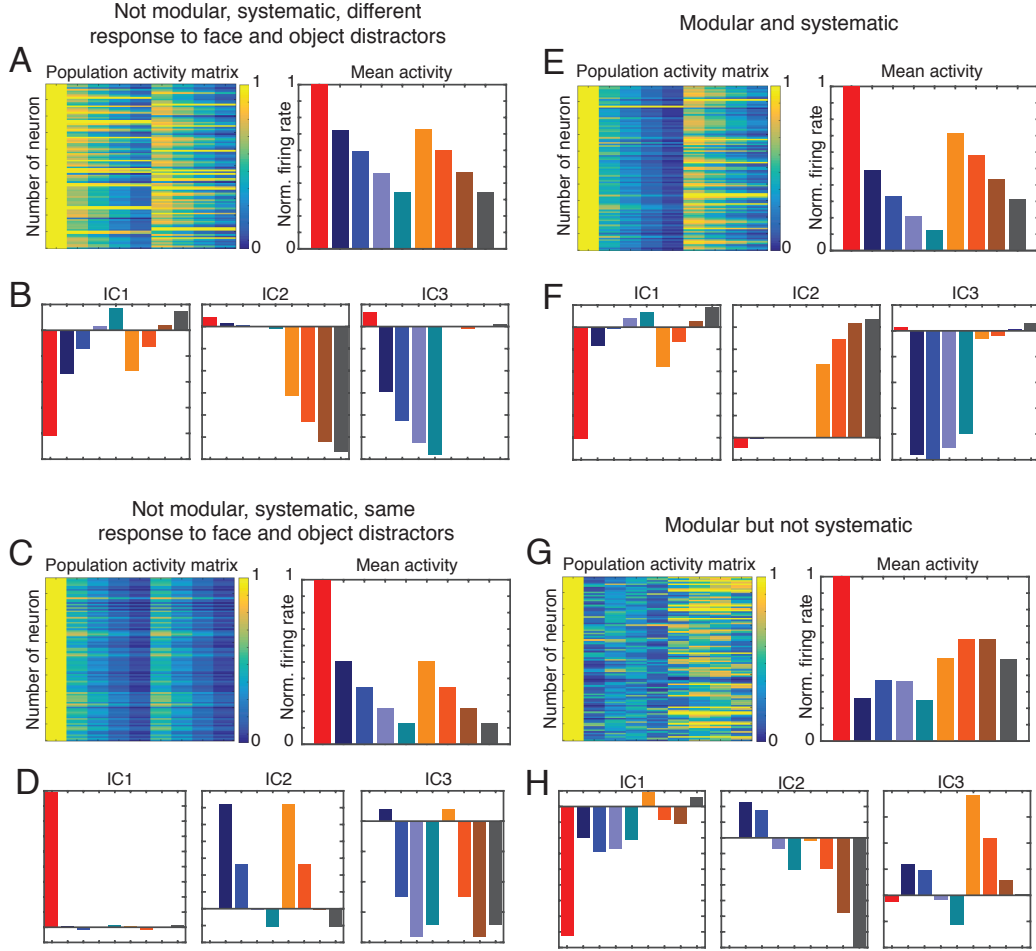


Figure 5.8: Effect of modularity and systematicity. Neural population of $N = 100$ was simulated with (A, B, C, D) or without (E, F, G, H) modularity. Namely, the population in (A) was simulated with population activity (P_i^{dist}) taken from the same distribution to produce responses to both face distractor and object distractor conditions. However, because two values for face and object conditions were taken randomly, the response to face and object distractor conditions (2-5 and 6-9 columns, respectively) were different. The population average across neurons is shown right. (B) Result of ICA applied to the population activity in (A). Population shown in (C) was simulated same as (A), but exactly the same (P_i^{dist}) values were used for both face and object distractor conditions. Thus, columns 2-5 and 6-9 show the same response pattern. The population average across neurons is shown right, and ICA result is shown in (D). To simulate the population in (E), (P_i^{dist}) values for face and object distractor conditions were taken from two different Gaussian distribution with different mean. The population average across neurons is shown right, and ICA result is shown in (F). Population shown in (G) was simulated same as (E), but columns 2-5 and 6-9 are independently shuffled for each neuron such that the resulting response to different number of distracters have a unimodal tuning curve. The population average across neurons is shown right, and ICA result is shown in (H). Color code follows the Fig. 5.2.

6

Comparison to anterior face patch AL

6.1 Anterior face patch AL is located one position higher than MF in the cortical hierarchy

In addition to MF, we also recorded from one of the anterior face patch AL (anterior lateral) to compare neuronal responses to that of MF neurons. According to previous microstimulation and tracer injection studies [88, 167], AL is located at the lower lip of the STS in anterodorsal TE (TEad). AL neurons are known to connect to MF in a reciprocal manner (Fig. 6.1).

AL neurons are also tuned to the head orientation, but approximately 3/4 of the neurons have a particular type of selectivity coined as “mirror-symmetric tuning” [89]. As shown in Fig. 6.2, Mirror-symmetric tuned AL neuron respond to both one head orientation and opposite direction of face image (e.g. left and right, up and down. It does not have to be a precise mirror image: except

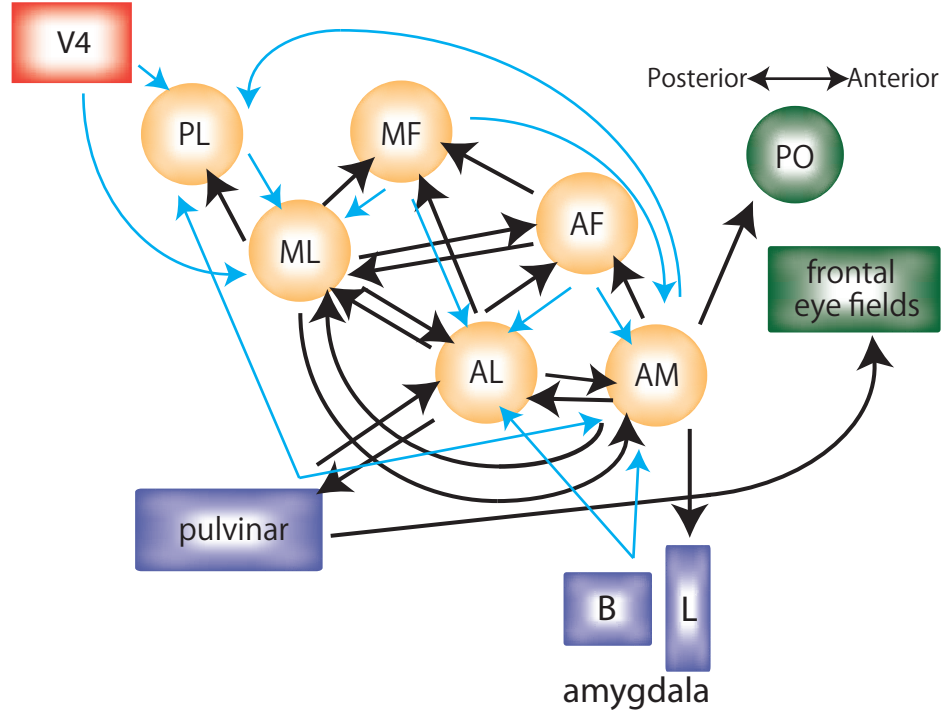


Figure 6.1: Face patch connectivity diagram. Arrows show directed connection between cortical regions. Black arrows are connection defined based on microstimulation study of Moeller et al. [88], and light blue arrows are defined based on tracer infection experiment of Grimaldi et al (conference presentation, unpublished [167]). Color code of brain regions are as follows: Red: early visual area, Yellow: temporal lobe, Green: frontal lobe, Blue: subcortical structures. Circles indicate face patches. V4: forth visual cortex, PL: posterior lateral, ML: medial lateral, MF: medial fundus, AL: anterior lateral, AF: anterior fundus, AM: anterior medial, PO: prefrontal orbital, B: basolateral amygdala, L: lateral amygdala.

few neurons tuned strongly to identity, neurons respond to broad range of visual images with the two preferred head orientations). Another noticeable physiological difference is its RF size, as expected from converging inputs to upstream neurons in the hierarchy [40].

Given that MF neurons could be a potential input source to the AL patch, AL neurons could be more robust against the distractors. Alternatively, AL neurons could show similar response property as MF, because (a) AL neurons may relay the response from MF, to show stronger reduction by distractors at the early phase, followed by more robust response against the distractors at the late phase. Or because (b) AL neurons might also receive inputs from other object selective patches (i.e. AL may receive inferences from object selective inputs similar to MF). Based on fMRI results, non-face object selective cortices are found in between the face selective cortices. Thus, even in the higher position in the hierarchy AL neurons might still receive (weaker) normalization from object patches and (stronger normalization from) face patches (however, the tracer study of Grimaldi et al. did not show direct evidence of inputs from object selective cortices).

We conducted recording from two AL patches of two monkeys (Monkey J and Monkey M, the former was also used in MF recordings). For the distractor category/numbers experiment and Prox/Ecc experiment, we collected 109 (48 from the monkey J and 61 from the monkey M) and 108 neurons (47 from the monkey J and 61 from the monkey M), respectively.

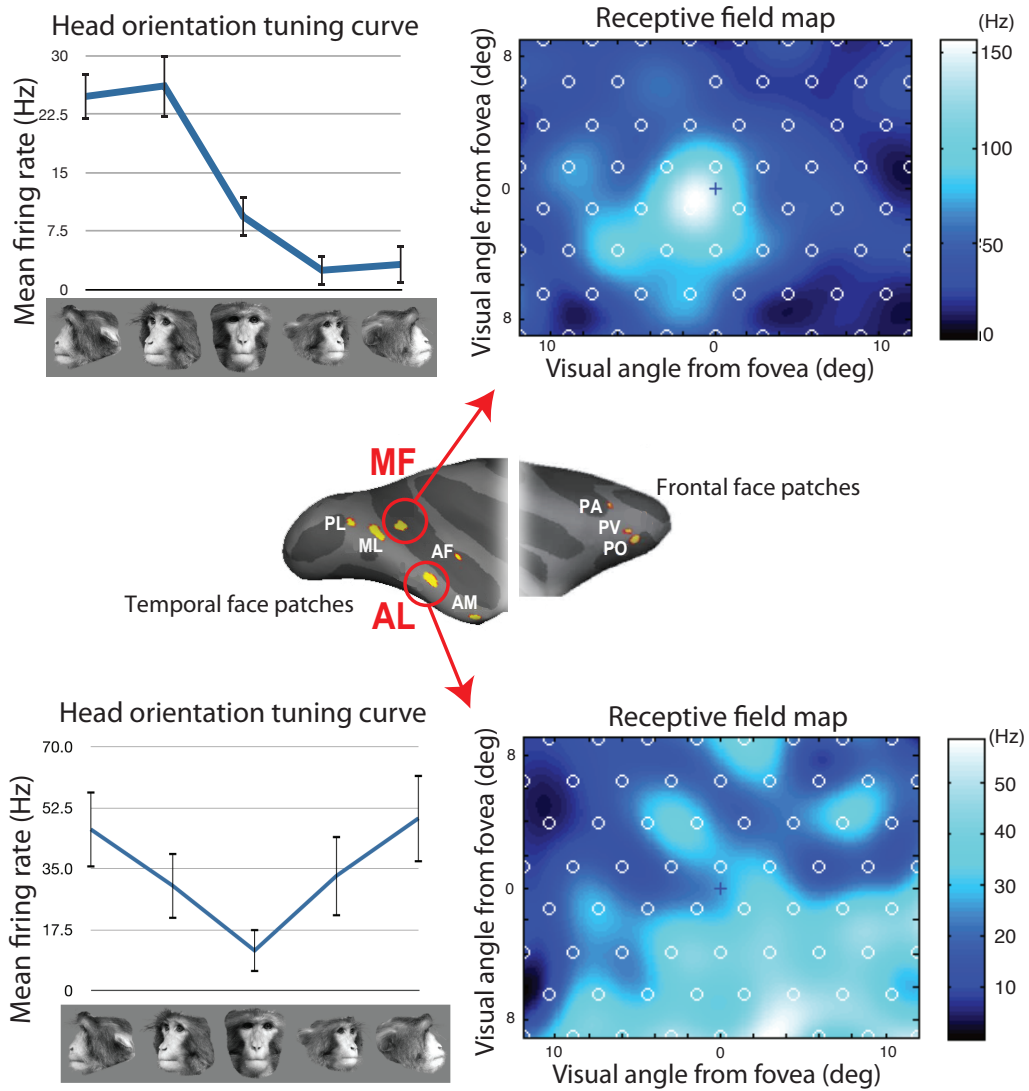


Figure 6.2: Comparison of physiological property between MF and AL neurons. (Top) Most of MF neurons are tuned to one head orientation showing a unimodal tuning curve. Most of RFs are foveal and circular. (Bottom) Approximately 3/4 of the AL neurons show mirror symmetric tuning, responding to a head orientation and mirror-symmetric direction of the head orientation. RF can cover almost entire hemifield on the screen or bottom half shown as example, or even larger. However, we also found AL neurons with smaller foveal RFs.

6.2 AL neurons' responses were reduced by distractors similar to MF, but latency was shifted with distractor numbers

As expected from the later position in the hierarchy, AL neurons' response latency (time from the stimulus onset to the response onset. Do not confuse with onset-to-peak latency, which I introduce later in this chapter) distribution was significantly different from that of MF neurons (Two-sample Kolmogorov-Smirnov test, $p=0.00165 < 0.05$). Specifically, neurons with very short latencies found in MF were not seen in AL neurons. This made the AL latency distribution less skewed than MF distribution, although the median value is only slightly increased (84 ms for MF distribution, and 89 ms for AL distribution) Fig. 6.3 shows the distribution of the response latency. As in the MF study, the response latency was calculated based on neuron's response to its preferred face: the response onset is the time when the trial-average response crossed 3 standard deviation of spontaneous activity (for details, see Appendix A).

We calculated the AL population average firing rate in the same way as for MF neurons. Specifically, response latencies were subtracted before averaging in order to align the response peak (response onset latency ranged from 56 ms to 254 ms). Note that the relative response lag across stimuli were preserved: response latency was defined per neuron, not per stimulus condition. Fig. 6.4 shows the result. As seen in the figure, the face distractors and object distractors still suppressed (or normalized) the neural response, quite comparable to what we found with MF neurons. Two-way ANOVA (factor 1: distractor number, factor 2: distractor category) verified significant reduction by distractor number difference both at the early phase ($F(4, 975) = 39.33, p = 1.46 \cdot$

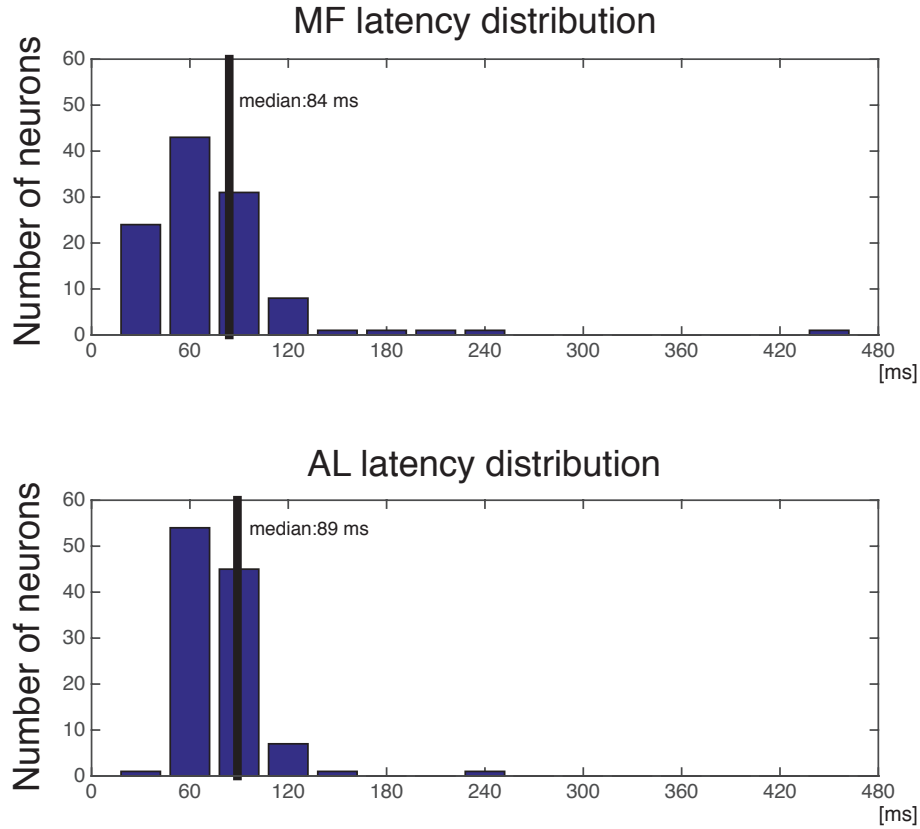


Figure 6.3: Response onset latency comparison between MF and AL. (top) Latency distribution from 111 MF neurons. The black vertical line is showing the median value: 84 ms from the stimulus onset. (bottom) Latency distribution from 109 AL neurons. The median value is 89 ms from the stimulus onset. Two-sample Kormogorov-Smirnov test rejected the null hypothesis that the two distributions were taken from the same distributions at 0.05 significance level.

$10^{-30} < 0.001$) and at the late phase ($F(4, 975) = 7.57, p = 5.21 \cdot 10^{-6} < 0.001$). Also, face distractor, rather than object distractor, suppressed neural firing rate significantly at the early phase ($F(1, 975) = 33.44, p = 9.87 \cdot 10^{-9} < 0.001$) but not at the late phase ($F(1, 975) = 3.62, p = 0.0575 > 0.05$). For the detail of Tukey post-hoc test, see Appendix J. See Table 6.1 for mean firing rate and SEM.

Table 6.1: Cat/Num experiment, AL population average response

	Early		Late	
	Face	Obj	Face,	Obj
0 dist.	1 ± 0.0509	1 ± 0.0509	1 ± 0.0409	1 ± 0.0409
1 dist.	0.747 ± 0.0430	0.861 ± 0.0520	0.839 ± 0.0381	0.859 ± 0.0400
2 dist.	0.549 ± 0.0421	0.710 ± 0.0470	0.716 ± 0.0372	0.809 ± 0.0417
4 dist.	0.391 ± 0.0376	0.630 ± 0.0524	0.636 ± 0.0409	0.728 ± 0.0421
8 dist.	0.307 ± 0.0380	0.538 ± 0.0514	0.564 ± 0.0398	0.722 ± 0.0416

There were two noticeable difference between MF and AL neurons. Firstly, AL neurons showed a systematic response peak shift with an increasing number of distractors (Fig. 6.4, top). On average MF neurons did not show such peak shifts (See Fig. 3.1), although few neurons did show small latency shift as distractor number increased. Secondly, peak magnitude of the distractor only conditions (Fig. 6.4, bottom) showed smaller response magnitude compared to MF responses. Especially the responses to the object distractors were very small and even showed a suppressive effect. This indicates that the face patch AL, later in the cortical hierarchy, shows stronger face selectivity, which was independently quantified by Meyers et al. (unpublished data, SfN Nanosymposium presentation [168]). However, as Fig. 6.4 shows, even with the reduced effect of object distractors, the neural response still showed response reduction by object distractors.

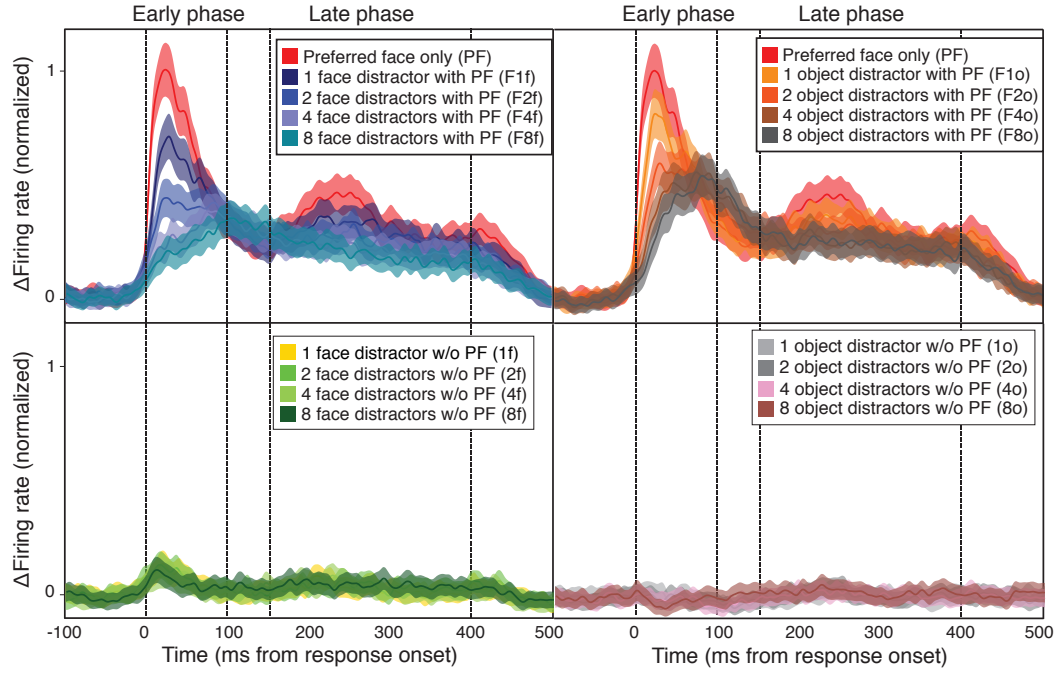


Figure 6.4: AL population average firing rate to Cat/Num experiment. Population average was calculated over 109 recorded neurons. Response onset latency of each neuron was subtracted to align the response phases. (Top left) Preferred face and face distractor conditions. (Top right) Preferred face and object distractor conditions. (Bottom left) Face distractor without preferred face condition. (Bottom right) Object distractor without preferred face condition. Color code follows Fig. 3.2.

To quantify the shift of response peak, we calculated the time of the maximum value of the response both in the early and the late phase. Specifically, trial average of the responses of each neuron was smoothed by a Gaussian kernel with 20 ms width (standard deviation of the Gaussian kernel) to reduce jitter affecting the max calculation. The result is shown in Fig. 6.5. Hereafter I call the quantified maximum time as onset-to-peak latency to prevent confusion with onset response latency (Response onset latency was subtracted from the response: 0 ms indicates the first instance when the response crossed 3 standard deviation of the spontaneous activity). In the early phase, 3-way ANOVA (factor1: ML vs. AL, factor2: distractor numbers, factor3: distractor category) showed that the AL onset-to-peak latency was significantly longer than that of MF ($F(1, 1754) = 100.92, p = 0.0000... < 0.001$), and increasing distractor number also significantly lengthen the onset-peak latency ($F(3, 1754) = 25.96, p = 0.0000... < 0.001$). In the late phase, AL onset-to-peak latency was significantly longer too ($F(1, 1754) = 8.18, p = 0.0043 < 0.05$), but latency increase due to distractor numbers did not reach the significance ($F(3, 1754) = 2.55, p = 0.0543 > 0.05$). Neither early nor late phase showed significant onset-to-peak latency difference between face and object distractors ($F(1, 1754) = 0.29, p = 0.591$ and $F(1, 1754) = 1.84, p = 0.175$, respectively).

We also calculated the latency-subtracted population average of Prox/Ecc variation experiment (Fig. 6.6). The result looked quite similar to that of MF response. For the RF center presentation, one-way ANOVA testing mean difference showed significance both at the early phase ($p = 3.67 \cdot 10^{-17} < 0.001$) and at the late phase ($p = 1.42 \cdot 10^{-11} < 0.001$). Also the RF periphery presentation condition reached significance both at the early phase ($p = 5.49 \cdot 10^{-9} < 0.001$) and at the late phase ($p = 1.57 \cdot 10^{-11} < 0.001$). For the details of comparison between pairs of stimulus condition, see Appendix J. For mean firing rates and

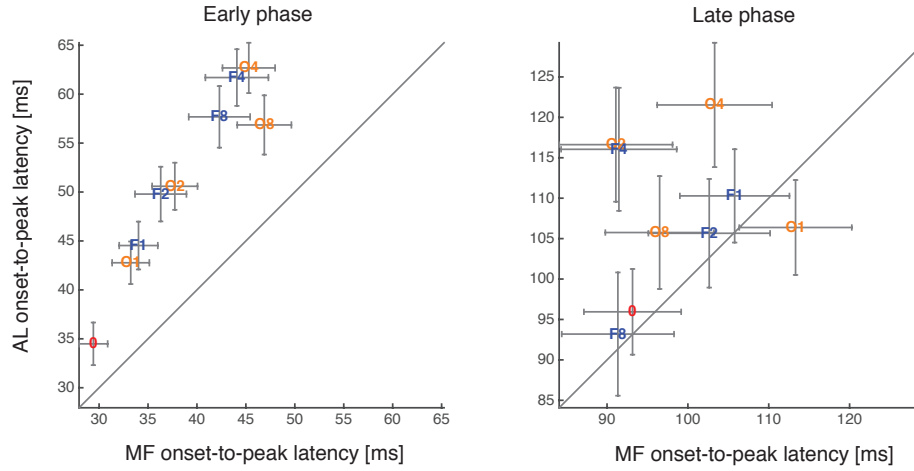


Figure 6.5: Response onset-to-peak time, compared between AL and MF neurons. The duration between response peak time (calculated for each stimulus condition) and the response onset time (calculated using the response to one preferred face without distractors). Note that the time was calculated from response onset, not stimulus onset (AL neurons are located later than MF in the cortical hierarchy, thus AL neurons respond later than MF neurons. By subtracting the latency we could have direct comparison only on peak shift due to the stimuli). The red “0” indicates the preferred face stimulus without distractors. F1, F2, F4 and F8 indicate 1, 2, 4 or 8 face distractors with preferred face. O1, O2, O4 and O8 indicate 1, 2, 4 or 8 object distractors with preferred face. Early phase showed significant onset-to-peak time shift across regions and across distractor numbers. The late phase showed significant onset-to-peak time shift only. Neither phase showed time shift depending on the object category (for details, see the main text). Error bars show standard error of the mean.

SEM, see Table 6.2.

Table 6.2: Prox/Ecc experiment, AL population average response

	Early		Late	
	Center	Periphery	Center,	Periphery
0 dist.	1 ± 0.0555	1 ± 0.0940	1 ± 0.0433	1 ± 0.0609
L dist.	0.695 ± 0.0483	0.405 ± 0.0722	0.627 ± 0.0358	0.465 ± 0.0411
M dist.	0.594 ± 0.0447	0.452 ± 0.0745	0.542 ± 0.0337	0.371 ± 0.0419
S dist.	0.396 ± 0.0445	0.423 ± 0.0712	0.428 ± 0.0352	0.319 ± 0.0416

Thus, so far we find response difference only in distractor numbers (which induced latency shifts), out of stimulus parameters we used: distractor numbers, category and proximity. In order to test whether the pattern of population response encodes stimulus information in the high dimensional space and whether the Divisive Normalization can explain the experimental findings (other than latency shift), we conducted the further analyses in the next section.

6.3 AL neural population preserved stimulus information in the representation space, revealed by ICA

As highlighted in the Chapter 4, the heterogeneous response pattern in the population activity could represent detailed information such as distractor number, category and proximity. We studied if AL neural population also contained information in response patterns by using dimension reduction followed by ICA. Fig. 6.7 showed quite similar pattern to the MF result shown in Fig. 4.5, representing distractor number, category and proximity as different

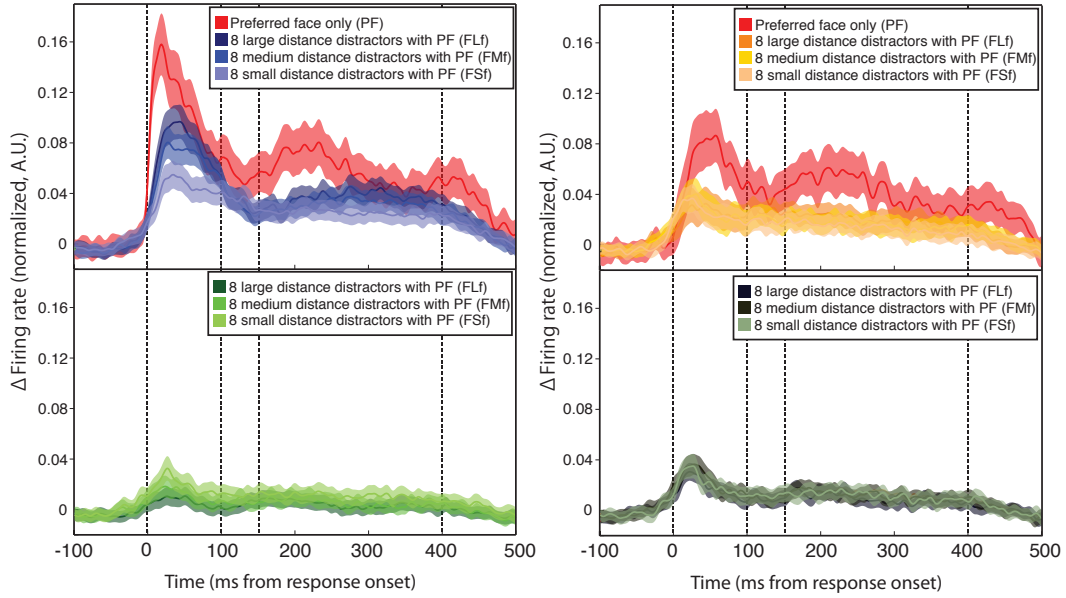


Figure 6.6: AL population average firing rate to Prox/Ecc variation experiment. Population average was calculated over 108 neurons. Response onset latency of each neuron was subtracted to align the response phases. (Top left) Preferred face with or without distractors of different proximities, RF center presentation. (Top right) Preferred face with or without distractors of different proximities, RF periphery presentation. (Bottom left) distractors of different proximities without the preferred face, RF center presentation. (Bottom right) distractors of different proximities without the preferred face, RF periphery presentation. Color code follows Fig. 3.5.

ICs (for details, see the Chapter 4).

6.4 Divisive Normalization explained the responses of AL neurons

As shown in the Fig. 6.2, RF of AL neurons have different structure than MF neurons. Since our computational model is highly relying on RF value, we wanted to test if the Divisive Normalization model still can explain the neurophysiological property of the AL neurons. As with MF analysis, we started from the pure prediction model - an equation without any free parameters. Fig. 6.8 shows the result at the early phase of the distractor category/number variation experiment. The pure prediction model followed the global response pattern: face distractor suppressed the neural response more than the object distractors, with increasing suppression as the number of distractor increases. However, one noticeable difference was the error at the object distractors without preferred face condition. Here, the actual average response shows suppression below the spontaneous activity level, whereas the model prediction is above the spontaneous activity. As a consequence, the pure prediction model expected larger normalization by the object distractors resulting in slight underestimation in the object distractors and preferred face conditions. Potential reason for this apparent discrepancy is the difference in the stimulus presentation duration and inter-stimulus interval (ISI). In the distractor category/number variation experiment we presented stimulus for 400 ms with 200 ms ISI, whereas in the RF measurement we presented for 200 ms with no inter-stimulus interval. Difference in response duration and ISI are known to affect response shape in visual, somatosensory and auditory cortex [169, 170, 171] and potentially information transfer efficiency, content and

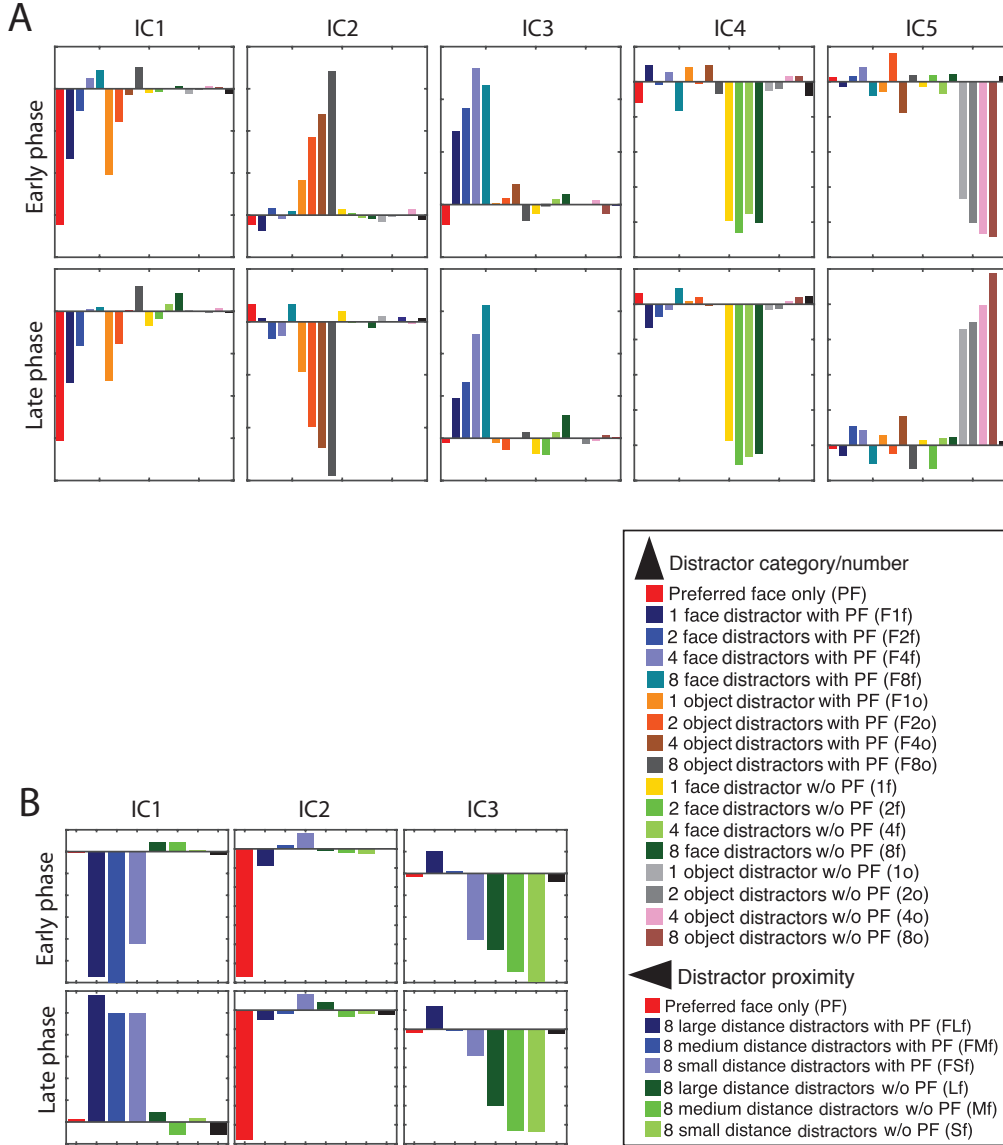


Figure 6.7: ICA on AL experimental results. The recorded neural responses were averaged within the early phase and the late phase to create population response matrices. The first dimension of the matrices was reduced to 5 and 3 for the (A) Cat/Num experiment and (B) Prox/Ecc experiment, respectively. ICA was applied after the dimension reduction to find independent components with minimal mutual information.

tuning curve ([169] and personal communication with Wilbert Zarco).

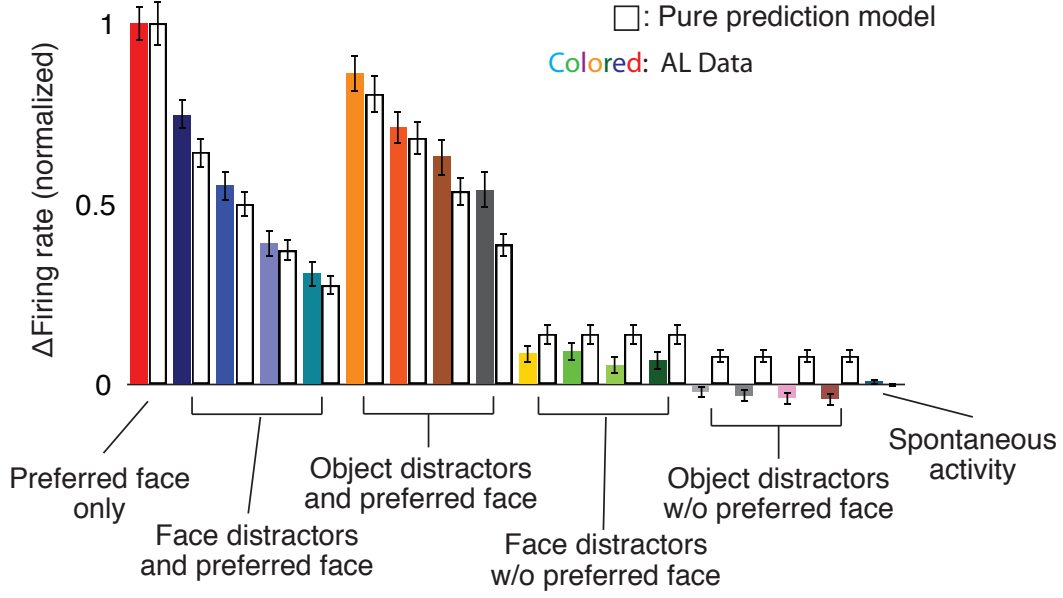


Figure 6.8: Pure prediction model on AL distractor category/number variation experiment. The mean firing rate calculated at the early phase, across trials were used for data. Same as modeling on MF data, Pure prediction model were calculated for each neuron using values from the measured RFs. Data and model from recorded 109 neurons were averaged to be shown as the bar graphs. Error bars show the standard error of the mean.

Next we introduced the post-diction model with an exponent n as a free parameter, as with MF analysis, to explain the rest of the conditions. The post-diction model could reproduced the remaining late phase and Prox/Ecc experiment, as shown in the Fig. 6.9. Finally, we used the model output of all recorded neurons for ICA to study the representation in the high dimensional representation space. The model output showed qualitatively same representation patterns as experimental data shown in Fig. 6.7, except the IC3 from the Prox/Ecc experiment at the late phase (Fig. 6.7C, bottom right). The late phase ICs could not represent the distractors without preferred face. It remains to be studied further whether this discrepancy was due to the model limitation, or simply due to the small explanatory power of the IC3: In the al-

gorithm we used, ICs were calculated in order of its explanatory power, which was equivalent to the explained variance in PCA. If representation of particular stimulus condition is stronger, it might appear in IC to exclude other conditions. Indeed, when the numbers of ICs were increased to 5, the distractor-only condition was represented similarly to the IC3 in Fig. 6.7.

6.5 Summary

In conclusion, AL neurons showed qualitatively similar response to that of MF neurons both in Cat/Num and Prox/Ecc experiments. Divisive normalization model also explained the response patterns across neurons, and ICA could represent independent information such as distractor category, number and eccentricity in different independent component as well. However, there were differences from MF, such as the latency shift due to increasing number of distractors and RF size (and, as a consequence, responses to multiple preferred faces). These differences are beyond this thesis's scope, but need to be investigated further in future experiments.

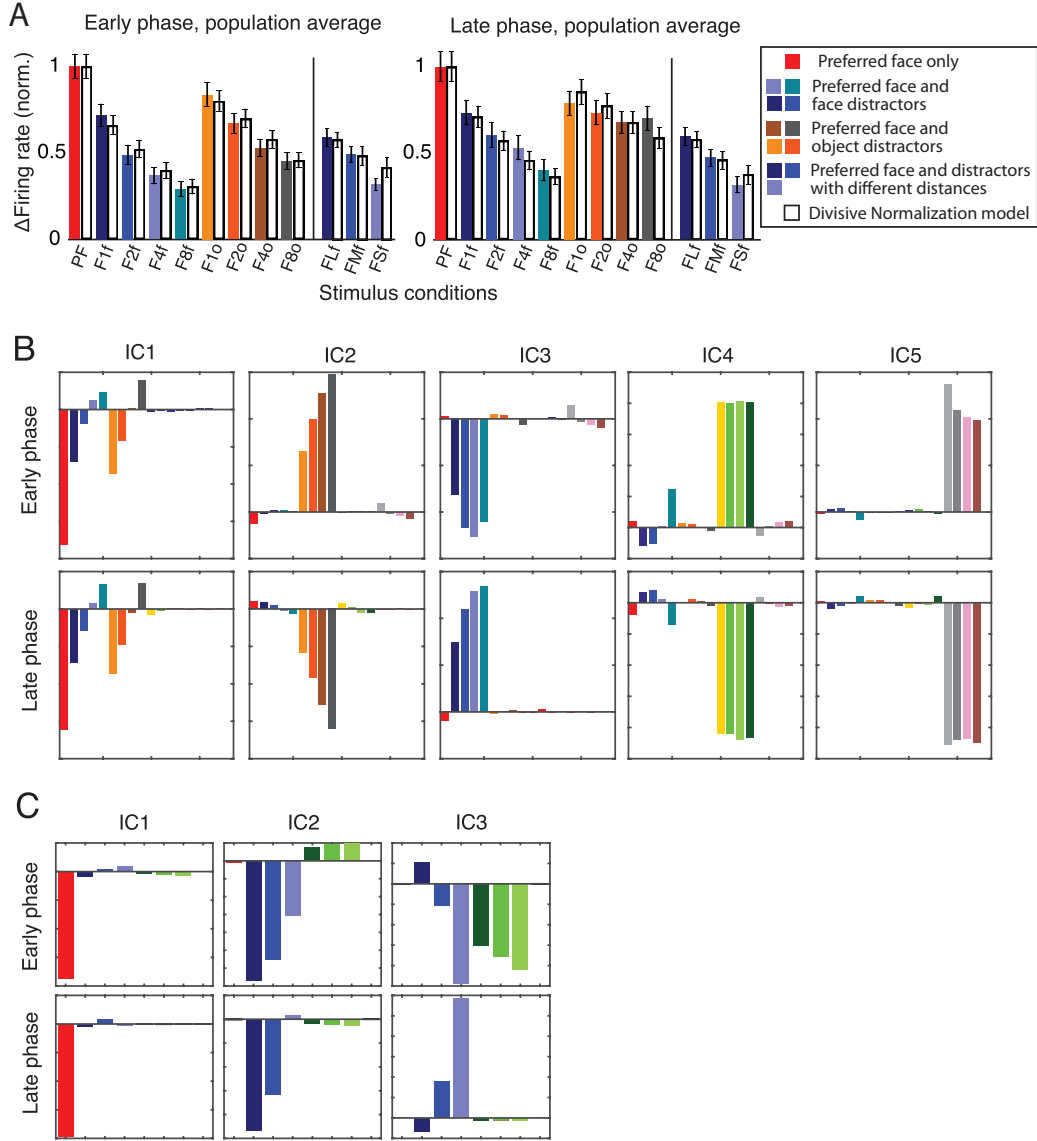


Figure 6.9: Postdiction model on AL experiment and population readout. Color code follows Fig. 5.4. (A) Comparison of average firing rate and average model output across neurons. Right: early phase, left: late phase. (B) ICA results for Cat/Num experiment. (C) ICA results for the Prox/Ecc experiment. RF center presentation result is shown.

7

Discussion

7.1 Heterogeneity and systematicity with Divisive Normalization represent multiple stimuli in population activity

With fMRI-targeted electrophysiology, we systematically varied distractor number, distractor category and distractor proximity to observe overall systematic reduction of mean firing rates in most complex stimulus conditions. However, even under the reduced firing rate, stimulus information was well preserved and discriminable in the high-dimensional representational space. In Chapter 4, we described how heterogeneity of neural population enables the representation in the population activity pattern. In Chapter 5, we showed that systematic response according to divisive normalization can enable the representation. We used divisive normalization to explain the neural responses. In terms of divisive normalization, response reduction due to increasing number of distractors was explained as follows: As the number of presented stimuli increases, the stimuli evoke large population activity in the neural population, resulting in

larger normalization power. Because the normalizing power is proportional to the number of stimuli, a neuron responds systematically to increasing number of stimuli. For example, as the distractor number increases from 0, 1, 2, 4 to 8, neural response decreases systematically. This systematicity is an important factor to have separable stimulus representation in population activity. Fig. 7.1 illustrates the importance of systematicity.

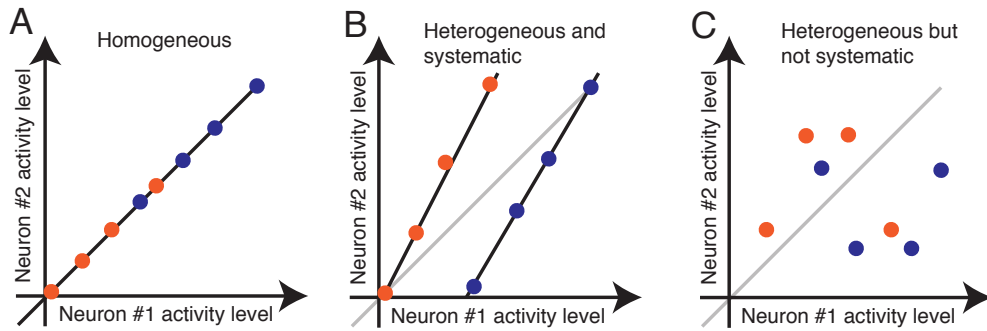


Figure 7.1: Systematic and heterogeneous responses discriminate stimulus groups. Hypothetical neural population consists of two neurons. Blue and orange dots belong to different stimulus groups (e.g. face distractor and object distractor condition). (A) Homogeneous neural response. Dots from two groups are intermingled at the diagonal. (B) Heterogeneous and systematic neural response. The two neurons respond differently yet systematic, grouping same stimulus condition. Two black lines indicate hypothetical representation axes created by ICA. (C) Heterogeneous but non-systematic responses. Stimulus conditions are intermingled and cannot be separated linearly.

Each panel of Fig. 7.1 shows a hypothetical neural population activities consist of two neurons. Fig. 7.1A is a homogeneous neural population. As we saw in Chapter 4, a homogeneous neural population cannot have linearly separable representations of different stimulus groups, shown in blue and orange dots in the figure. Fig. 7.1B and C show heterogeneous neural populations, but neurons in B respond systematically to each stimulus group (imagine the four blue and orange dots representing 1, 2, 4 and 8 distractor condition and

neurons change response systematically). Because of the systematicity, the blue and orange stimulus group are linearly separable¹ and ICA can find representation axes highlighted by black lines². However, neurons in Fig. 7.1C lack systematicity: in this example, the neuron #1 is responding monotonically (ramping tuning curve, e.g. reducing response as distractor number increases from 1 to 8) whereas the neuron #2 is tuned to one particular stimulus condition (unimodal Gaussian tuning curve, e.g. highest response to 2 and 4 distractor conditions). In this case, the blue and orange dots cannot be separated by a line, unless response magnitude evoked by the two stimulus are sufficiently different (in which case, even homogeneous population can discriminate the two groups). The simulation result in Fig. 5.8.G showed that disrupting systematicity impaired the ability of ICA to represent independent information in each axis, even though unimodality of the tuning curves were preserved. Thus, heterogeneity and systematicity are two important factors for MF and AL neurons to represent multiple stimuli in discriminable way. Each neuron responds to the stimulus in slightly different way (i.e. heterogeneous), but their responses are changing with distractor numbers as explained by divisive normalization (i.e. systematic), resulting in discriminable information representation in the space spanned by population activity.

¹Recently it was shown that linear separability is increased from lower to higher cortex in the cortical hierarchy, and the linearly separable representation was shown to be important to correctly guide animal behavior in complex tasks [172, 173]. Thus, although it remains elusive whether a brain can operate ICA-like decoding to read information out of face patch neurons, the brain may tend to have linearly separable representation which is beneficial for behavior.

²In terms of ICA, marginal distributions to the two new axes are non-Gaussian, or close to flat distribution in this example. For ICA and minimizing Gaussianity, see Chapter 4. Note that we used linear ICA.

7.2 Modular structure confers category discrimination

In the previous section, I discussed how heterogeneous and systematically responding neurons represent multiple stimuli. The systematic increase of the normalization power discriminated stimuli with different number of distractors, both in single neuron responses and in population readout. Moreover, distractors in different categories were also discriminated both in single neurons and in population readout. This is not trivial because as we saw in Chapter 3, distractors themselves elicit small and comparable responses regardless of whether distractors are faces or objects. Thus, initially we thought a single neuron could not discriminate “the preferred face and the non-preferred face”, and “the preferred face and the non-preferred object”, because the distractor, regardless of being face or object, elicit equally weak response to the neuron. No matter how the neuron compute these inputs (e.g. average or max), the output of the neuron to these two stimuli should be the same. However, we found that even a single neuron could discriminate these two stimuli, with larger response to the preferred face and object distractor. In Chapter 5, we showed that heterogeneity is necessary to discriminate category in high-dimensional space spanned by population, but in order to discriminate in single cell level or in population average response, modularity is required. The face selective neurons are segregated in patch-like module structure in the temporal lobe [67, 71, 70]. Because neurons with similar selectivity are clustered as a modular structure, they can have a large activity to stimuli containing their preferred category (in this case, face). In terms of divisive normalization, presenting face distractors induces large population activity to give a larger normalization power and a larger weight for the face distractors compared to

object distractors. This allows single neurons to discriminate two input groups of different categories whose magnitude are equivalent to single neurons. This discriminability cannot be achieved if neurons are distributed and connected randomly across the temporal lobe. It has been an open question why the neurons have to be clustered (although few hypotheses were proposed such as efficient wiring: see [174, 175]), but the current study gives an important hint to the question: The face and object selective neurons in the temporal lobe might form modular structures in order to confer category discriminability to single neurons.

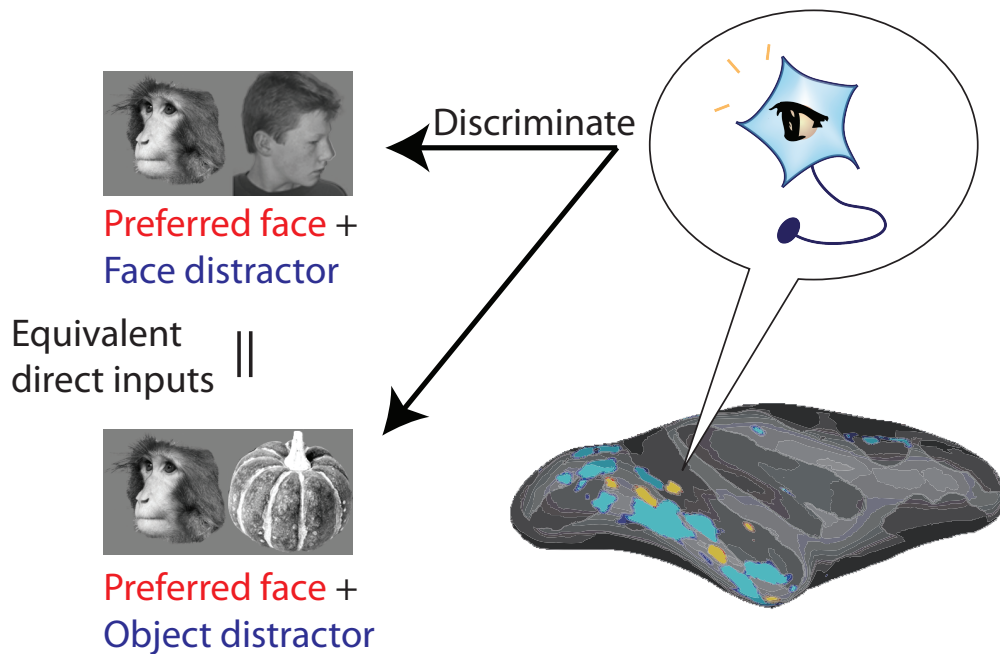


Figure 7.2: Modular structure confers category-discriminability. Two different stimuli, the preferred face combined with a face distractor and the preferred face combined with an object distractor provide equivalent direct inputs to a neuron. Therefore simple summation, average or max operation cannot discriminate these two. However, due to normalization specific to stimulus category a single visual neuron is capable of discriminating these two seemingly equivalent inputs. That is because temporal lobe object selective neurons are segregated into patches and connected each other to provide a category-specific normalization power to represent and discriminate complex stimuli.

7.3 Effect of modularity, heterogeneity and systematicity on representation space

In order to have an intuitive understanding of the effects of modularity, heterogeneity and systematicity, we performed PCA and ICA on simulated population activity created with Equation 5.5 and Equation 5.6. Fig. 7.3 explains an example visualization of PCs and ICs in two-dimensional space. Two component's scores are plotted as a two-dimensional scatter plot to have an intuition how the stimulus conditions are separated from each other. As in Chapter 5, 9 stimulus conditions (preferred face only, 1, 2, 4 or 8 face distractors with a preferred face, 1, 2, 4 or 8 object distractors with a preferred face. See figure legend in Fig. 7.3) from Cat/Num experiment are used. The population activity used in Fig. 7.3 has modularity, heterogeneity and systematicity. Thus, resulting PCA and ICA results look very similar to that of recorded data. When PC1 and PC2 are plotted (Fig. 7.3, top right panel) together, it is clearer that face distractor conditions and object distractor conditions are separated in two clusters. Namely, preferred face (PF) is on top right corner, and as face distractor number increases from 1, 2, 4 to 8, face distractor conditions (F1f, F2f, F4f, F8f) move away from PF (toward top left). Instead, when object distractor number is increased, object distractor conditions (F1o, F2o, F4o, F8o) also move away from PF but in different direction (Fig. 7.3, toward bottom right). As described in Fig. 7.1, these systematic distribution of stimulus conditions can be captured well by ICA, as shown in bottom panels in Fig. 7.3. With the next few simulations, I explain how each of the modularity, heterogeneity and systematicity contributes to what aspect of this stimulus representation in 2D space.

Firstly, a neurons that give an uniform response to all stimulus conditions

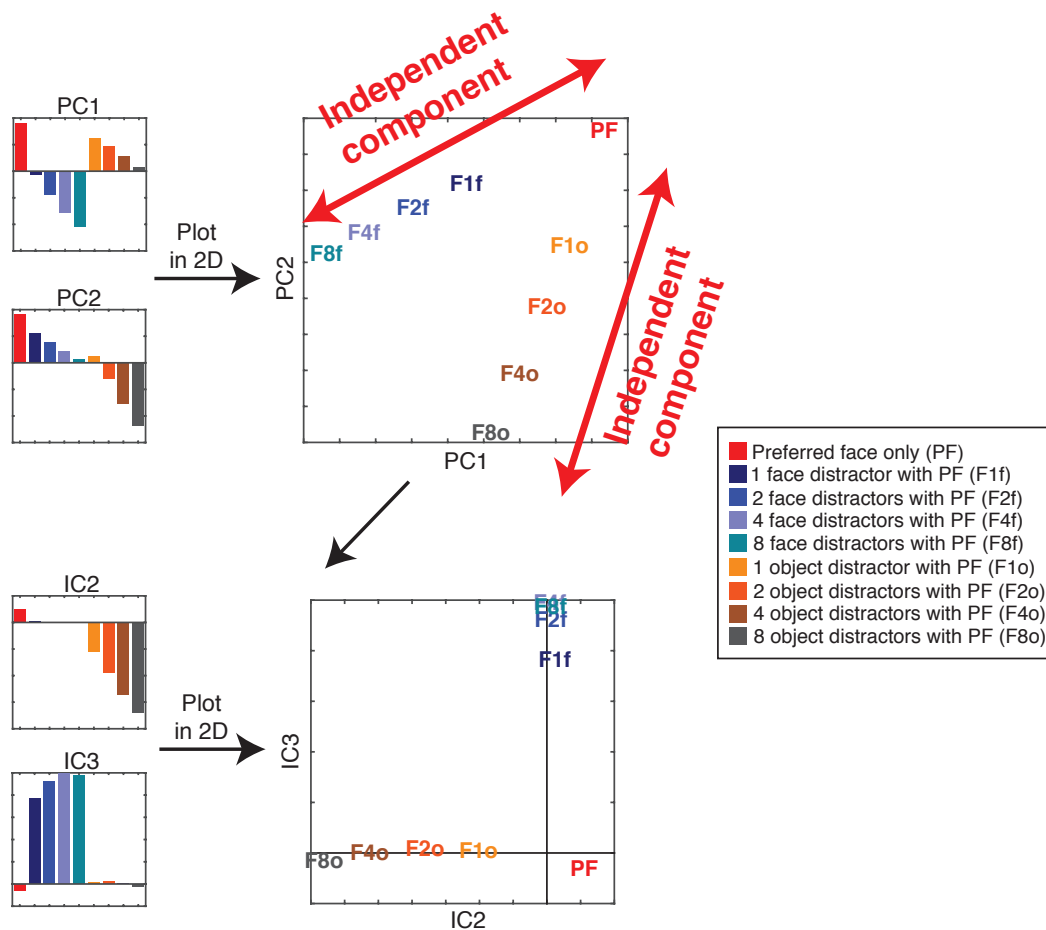


Figure 7.3: Schematic drawing showing how ICA finds independent representations from dimension-reduced data. Top: Score of two principal components, PC1 and PC2 are visualized in two-dimensional plot (PC space). Bottom: ICA extracts two independent components, IC2 and IC3 from the PC space, also visualized in two-dimensional plot (IC space). Note that the selected PCs and ICs are arbitrary, but the same logic can be applied to any PCs and ICs.

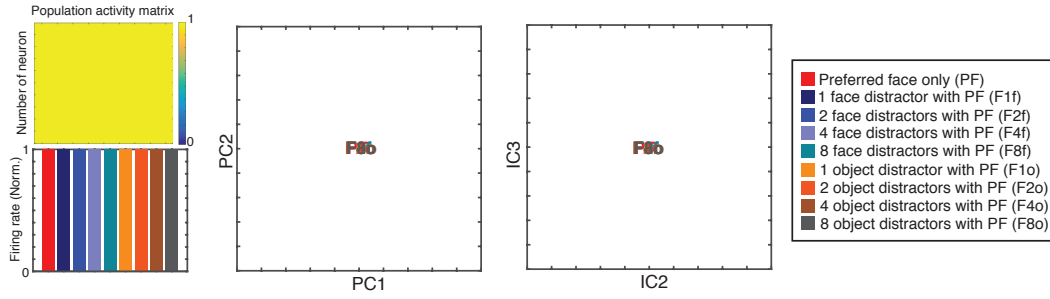


Figure 7.4: PCA and ICA on a complete homogeneous response matrix. Left: simulated population activity matrix and average firing rates across neurons. Middle: PC1 and PC2 from PCA on the population activity matrix. Right: IC2 and IC3 on the population activity matrix.

are considered. Regardless of stimulus condition, any neuron in this population gives response “1” when one of the stimuli is presented. Fig. 7.4 shows the population activity matrix (left, top. column: stimulus condition, row: neuron), average firing rate across neurons (left, bottom) and PCA (middle) and ICA (right) results. Mean firing rate is always 1 across stimulus conditions, and neither PCA nor ICA can separate stimulus conditions in two-dimensional space.

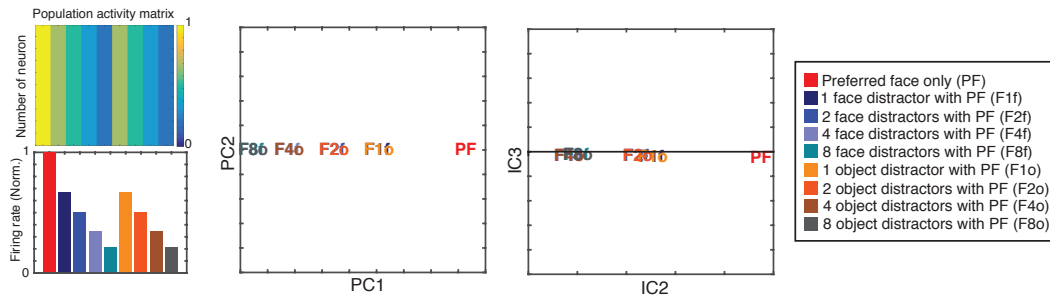


Figure 7.5: PCA and ICA on a systematic and homogeneous response matrix. Left: simulated population activity matrix and average firing rates across neurons. Middle: PC1 and PC2 from PCA on the population activity matrix. Right: IC2 and IC3 on the population activity matrix.

Next, systematicity is added to modulate firing rates according to increas-

ing distractor number. Namely, responses are generated according to divisive normalization (Equation 5.5 and Equation 5.6) to create systematic response reduction as distractor number increases (Fig. 7.5). In the following simulations, the single neuron responses S_i^{pref} , $S_i^{\text{dist,face}}$ and $S_i^{\text{dist,object}}$ and the population response to a preferred face $P_i^{\text{dist,face}}$ are fixed to be the same as the simulation in Chapter 5. In the population of Fig. 7.5, the population responses to a face distractor, $P_i^{\text{dist,face}}$ and to an object distractor, $P_i^{\text{dist,object}}$ have exactly the same values and fixed across all neurons. Thus, each neuron reduces responses as distractor number increases, but distractor category does not modulate the responses. Also $P_i^{\text{dist,face}}$ and $P_i^{\text{dist,object}}$ are the same across neurons. As shown in Fig. 7.5, average firing rates are the same for face distractor and object distractor, if the distractor number is the same (Fig. 7.5, left). However, because firing rates decrease with increasing number of distractors, one can tell the number of distractors by measuring the mean firing rate. Similarly, in PC and IC space, stimulus conditions with different distractor number are separated, although face and object distractor conditions are overlapped. Thus, systematic responses predicted by divisive normalization can separate stimulus conditions according to distractor number.

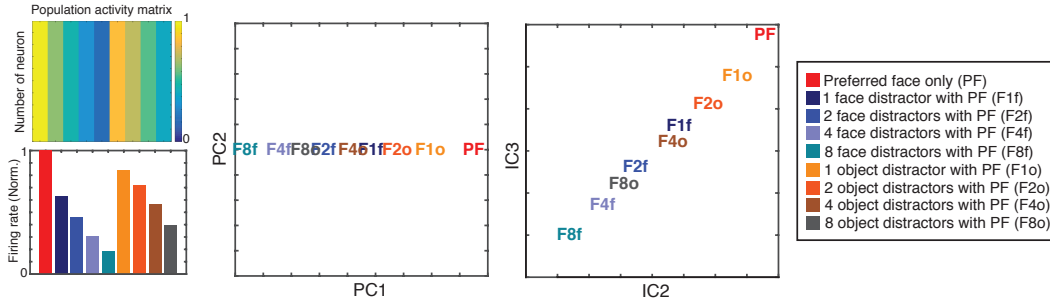


Figure 7.6: PCA and ICA on a systematic, modular and homogeneous response matrix. Left: simulated population activity matrix and average firing rates across neurons. Middle: PC1 and PC2 from PCA on the population activity matrix. Right: IC2 and IC3 on the population activity matrix.

When a modularity is added in addition to systematicity, firing rates of single neurons and population average could discriminate distractor category by providing larger firing rate to object distractor condition than face distractor condition (Fig. 7.6, left). In order to include modularity in the simulation, $P_i^{\text{dist,face}}$ is set to a larger value than $P_i^{\text{dist,object}}$ to have larger normalization (i.e. smaller firing rate) when face distractors are presented. Similar to the mean firing rate, the result of PCA (Fig. 7.6) shows separation between face distractor conditions and object distractor conditions. Thus, modularity can modify mean firing rate to discriminate distractor category (as explained in Chapter 5, PC1 is closely related to the mean firing rate). However, conditions are only separated in one dimension (PC1 direction) and PC2 cannot separate conditions as in Figure 7.3. Accordingly, ICA cannot separate different distractor categories as independent components either: IC2 and IC3 values show very similar values, resulting in one dimensional separation same as PCA (Fig. 7.6, right) As shown in Chapter 5, this is because of homogeneous responses (i.e. all neurons give the same response patterns). In order to separate stimulus conditions in two dimensions, heterogeneity is introduced in the next simulation.

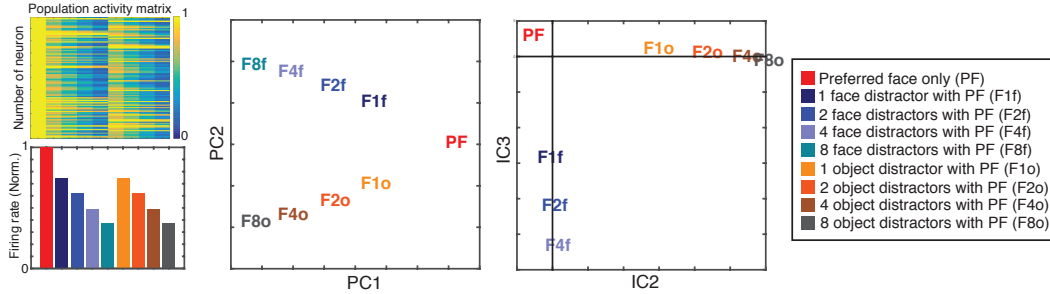


Figure 7.7: PCA and ICA on a systematic and heterogeneous response matrix. Left: simulated population activity matrix and average firing rates across neurons. Middle: PC1 and PC2 from PCA on the population activity matrix. Right: IC2 and IC3 on the population activity matrix.

To simulate heterogeneity, $P_i^{\text{dist,face}}$ and $P_i^{\text{dist,object}}$ are drawn from two Gaussian distributions with equal mean (i.e. no modularity, Fig. 7.7 or different mean (with modularity, Fig. 7.8). Without modularity, mean firing rate cannot discriminate distractor category if same number of distractors are used (Fig. 7.7, left), similar to Fig. 7.5. However, as shown in the middle panel of Fig. 7.7, PCA can separate two stimulus groups with different distractor category in PC2 direction (also see Fig. 5.8A, B). ICA captures these two separated stimulus groups to represent as two independent components as shown in the right panel of Fig. 7.7. Thus, in order to separate stimulus conditions in two dimensions, heterogeneity is required. This argument holds true for dimensions higher than two, highlighting importance of having heterogeneous response patterns across neurons.

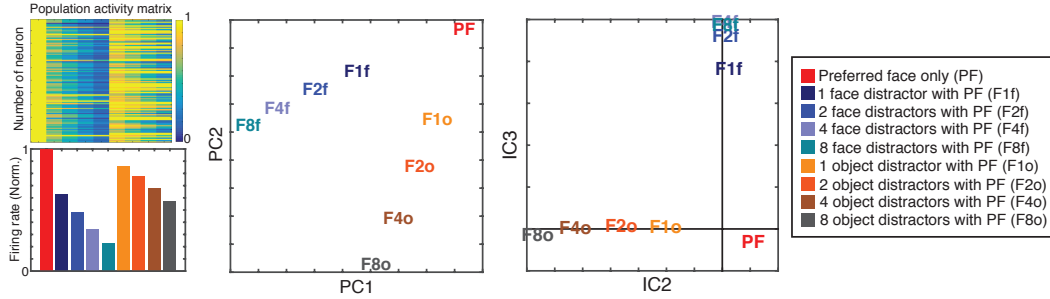
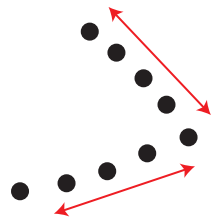
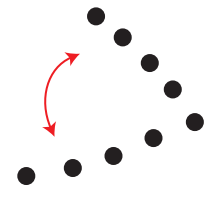
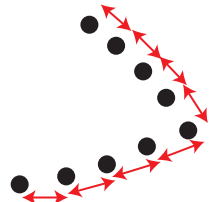


Figure 7.8: PCA and ICA on a systematic, modular heterogeneous response matrix. Left: simulated population activity matrix and average firing rates across neurons. Middle: PC1 and PC2 from PCA on the population activity matrix. Right: IC2 and IC3 on the population activity matrix.

Finally, modularity is added in addition to systematicity and heterogeneity. As a result, mean firing rate can discriminate face and object distractor conditions even when distractor numbers are the same (Fig. 7.8, left). Representation in PCA space looks rotated compared to that in Fig. 7.7, but ICA can capture the two stimulus groups to represent in two dimensional space, as shown in the right panel. Table 7.1 summarizes the role of modularity,

heterogeneity and systematically. Note that in all of above simulation I fixed S_i^{pref} , $S_i^{\text{dist,face}}$ and $S_i^{\text{dist,object}}$ while varied P_i^{pref} , $P_i^{\text{dist,face}}$ and $P_i^{\text{dist,object}}$, but the opposite led to the same conclusion. Namely, fixed P_i^{pref} , $P_i^{\text{dist,face}}$ and $P_i^{\text{dist,object}}$ with varying S_i^{pref} , $S_i^{\text{dist,face}}$ and $S_i^{\text{dist,object}}$ led to the same conclusion on modularity, heterogeneity and systematicity.

Table 7.1: Role of modularity, heterogeneity and systematicity.

	Description	Effect in 2D
Modularity	Separate different distractor category in mean firing rate	
Heterogeneity	Separate different category of stimuli in high dimensional space	
Systematicity	Separate different number of stimuli both in mean firing rate and high dimensional space	

7.4 Does a preferred face “pop-out”, or is it “crowded” by distractors?

Our result does not reject the existence of either hierarchical processing based on MAX-neuron or neurophysiological crowding effect (i.e. reduced discriminability by distractors). First, we indeed found some MAX-like neurons

(Fig. 3.4), which did not significantly reduce their responses in the presence of non-preferred distractors. This small number of MAX-like neurons might be contributing to invariant object recognition in primate vision. Second, larger firing rate reduction by face distractors compared to object distractors, and larger reduction by closer distractors are in line with human psychophysics of crowding effect, where intra-category distractors affect target identification more than inter-category distractors [135, 176]. However, we also found stimulus representation in population activity pattern in high-dimensional space, from which we could decode stimulus information such as distractor number, category or proximity. Thus, unless our perception depends solely on mean firing rate of the population activity, there is a possibility for a brain to decode and utilize information from neural activity. For testing whether an animal can utilize the information in the high-dimensional space, additional behavioral experiments will be required to see if an animal can identify the center stimulus in presence of distractors.

7.5 Pop-out effect and attentional confound

Perceptually, faces are known to automatically attract attention, or pop-out [123, 124]. This evidence raises a possible interpretation of our data that observed response reductions were due to the limitation of attentional resource: for example, multiple face distractors disperse attention over all stimuli including distractors, leading to overall response reduction. We do not think our result is due to attentional confound for the following reasons. First, stimulus numbers and category information was well preserved in the high-dimensional representational space, even when distractors were non-face objects. If attention works as a stimulus filter, this information should not be found. Second, if the animals were actively paying attention to faces to induce top-down at-

tention, attentional effect should be more prominent at the late phase [177]. This should lead to larger response reductions by face distractors at the late phase, but the observation was the opposite (response reduction compared to preferred face alone: 71, 45, 32 and 27% for 1, 2, 4 and 8 face distractors at the early phase, and 78, 64, 56, 44% for 1, 2, 4, 8 face distractors at the late phase, respectively. $F(1, 536) = 19.89, p = 1.00 \cdot 10^{-5}$). Third, the responses to distractors without the preferred face were comparable regardless of distractor numbers. If response reduction is due to dispersed attentional resource, larger number of distractors without the preferred face should show similar response reduction. Fourth, the Divisive normalization framework could explain the observation very well, indicating that the response reductions are mediated by normalization effect by the functionally segregated object selective neurons. From these results, we concluded that the observed response reductions are most likely caused by normalization.

As we discussed in the introduction, Divisive normalization and Biased competition model of attention can have a common formula. While the former describes normalization biased by stimulus contrast or category (e.g. face or object), the latter describes competition over attentional resources. Given that these observations can be explained by the same formula, it is possible that the normalization and attention involve a common circuit mechanism too. However, both of these models are phenomenological model lacking detailed definition of underlying circuit structure (e.g. number of nodes, degree distribution, direction of connection, etc.). Without a mechanistic model to analyze the circuit detail, it is difficult to conclude whether these two phenomena can be discussed in the same framework. Additionally, Divisive normalization is thought to appear during feedforward and lateral interaction (i.e. local process), while Biased competition is often (but not always) thought to be trig-

gered by top-down attention process. Whether these different causes can share a common circuit to shape neural responses will be a subject of future studies.

7.6 Divisive Normalization gives an unified account encompassing previous seemingly contradictory results

Our model explains and encompasses the two studies of Zoccolan et al. [117, 107], in which they used up to 3 visual stimuli at a time to study clutter tolerance. In the 2005 paper, they reported that responses of IT cells to multiple stimuli are the average of the responses to individually presented stimuli. In the 2007 paper however, they found some neurons showed stronger clutter tolerance (i.e. response larger than average), which correlated with neuron’s spatial tolerance (i.e. RF size) and anti-correlated with sparseness of selectivity. In our model, larger RF can encompass all presented stimuli to give larger direct drives, in order to compete with normalization power given by neighboring neurons. Indeed, the multiple preferred face condition showed that larger RF neurons had stronger clutter tolerance (i.e. less response reduction). This held true for broad selectivity of neurons. If a neuron can respond strongly to the distractors as well, both preferred face and distractors elicit large direct drives, again allow neurons to counteract against normalization.

Their earlier findings that neurons take average of the input may look contradictory with our results because, even though a face distractor and an object distractor elicit equally small response by themselves, a response to a preferred face and face distractors were closer to an average of responses to stimuli presented individually, while a response to a preferred face and object distractors elicited larger response than that predicted by an average. However,

Zoccolan et al. did not conduct recordings in a cortex with particular object selectivity, thus their neural population might not have strong preference to one of stimulus categories they used. Our experiment was conducted in a face-selective cortex, with face and non-face object being used. Because of this modular structure and properly selected stimulus database, we found category-dependent normalization. If we used a stimulus category such that the neural population does not have specific preference, our model gives an averaging response.

7.7 Divisive Normalization to prevent saturation and to have economical representation

We showed that Divisive Normalization model could explain the recorded data. One possible reason why neurons adopt this normalization is that it can prevent saturation. Although one of the easiest way to preserve all stimulus information might be summing up all of the multiple inputs a neuron receives, this strategy is hampered by the fact that neuron’s physiologically feasible firing rate range is limited. Unless a neuron can produce an infinite firing rate, neural activity will saturate quickly if all of the afferents are summed up. Instead, by normalizing inputs at each cortical region, neurons can maintain information by compressing representation within the physiologically feasible range.

Having low firing rates is also beneficial for economical representation, which was suggested by both experimental and theoretical studies. For example, Stokes et al. found very low population activity in the delay period of a working memory task, but with MDS (multidimensional scaling) they could find working memory representation in high dimensional space [160]. Also, Mongillo et al. proposed a model of working memory based on calcium buffer

(which can maintain short-term memory even with low firing rate), which is also beneficial for economical representation of information [178].

7.8 Representation at the late phase: robust against distractors yet maintaining representation

Related to economical representation discussed in the previous chapter, the recorded face selective neurons showed lower firing rate magnitude at the late phase³. We calculated the distractor robustness indices to find the neurons changed their robustness indices from averaging regime to robust regime toward the late phase. However, although neurons tend to “ignore” distractors at the late phase, multiple stimuli (including distractor stimuli without the preferred face) were represented and could be decoded from the population activity, as ICA and machine learning showed. In the previous chapter we discussed that this is due to the heterogeneity of population activity. Indeed, there were always a certain number of neurons in between the average and robust regime in the robustness indices distribution of the late phase (Fig. 5.5). At the late phase neurons are in general robust against their non-preferred stimuli, some of neurons even ignore the distractors almost completely, but there are always a certain number of neurons still responding to distractor, keeping all the information represented at the early phase. It was reported that IT neurons change response over time [49, 179]. Similarly, we found that the coding scheme changed from the early to the late phase, while keeping the representation almost intact. It would be interesting to know if this transition

³Note that the low firing rates at the late phase is not necessary due to Divisive Normalization *per se*, but could be an other mechanism such as adaptation.

changes under a particular behavioral demand (e.g. working memory task, object identification task) or during an information transfer to another cortical area (for example, see [173, 180])

7.9 Transition from MF to AL

Given that the neurons become robust against distractors at the late phase, and given that AL neurons are receiving afferents from MF neurons, one might expect that AL neurons are more robust against distractors than MF neurons. However, we found very similar response pattern among AL neurons: distractor still reduced mean firing rates. Why did AL neurons show similar response patterns and stimulus representation? One possible reason is because of other afferents from different cortices. AL neurons may receive inputs not only from MF neurons, but also from other non-face object selective cortices to have interference from distractor information. However, this is less likely in light of Grimaldi et al.'s result reporting no strong inputs to AL from object selective cortices [167]. Another possible reason is phasic relay: the late phase starts 250 ms after the response onset in our definition, which is much longer than the time required to synaptic transmission from MF to AL (Freiwald et al. reported 7 ms delay from MF (126 ms) to AL (133 ms), measured by local field potential [89].). Thus, the early phase of MF might be transmitted to AL to have similar early phase response, and similarly the late phase of MF might be transmitted to AL late phase. However, we observed increasing latency shift by increasing number of distractors in AL neurons (see Chapter 6), which cannot be explained by simple relay. An alternative possibility is common subcortical inputs. AL neurons receive inputs from basolateral amygdala, which is known to give feedback to broad regions of temporal lobe [167, 181]. Although it is not known if MF receives input from basolateral amygdala,

there is a certain possibility that the amygdala is providing common input. However, the amygdala response to face image is slow (around 300 ms from stimulus onset [58, 182]) and cannot explain responses in the early phase. In any case, the transition from MF to AL might be dedicated to have mirror-symmetric tuning [89], not to have robust representation against distractors. In fact, having completely robust representation means discarding information other than that of neuron’s preferred face. As a sensory cortex, temporal lobe may keep much information as possible and allow higher association cortex to select behaviorally relevant information as needed. Alternatively, the fact that we did not see a change in representation might be because the subject animals did not engage in any behavioral task. It would be interesting to study whether the representation and transition from MF to AL change under behavioral need to particular information. Also, the origin, mechanism and role of the latency shift observed in AL neurons remain to be studied. The latency shift could be due to the lag to reach the response threshold, but currently we do not have a suitable model to explain, nor do not have a good way to incorporate into the Divisive Normalization model. We speculate the Drift Diffusion model or Ornstein-Uhlenbeck process with drift strength depending on distractor number might work [183, 184, 185], but it should be pursued in future studies.

7.10 Concluding remark

Combining electrophysiology, fMRI and computer model, we studied how face selective neurons in the macaque temporal lobe represent multiple objects including faces and non-face objects. Although firing magnitude changes throughout the response phase, we found robust stimulus discriminability in the population level representation space in both the early and the late phase.

These results suggest that, aided by modularity, heterogeneity and systematic normalization, neural population can preserve robust representation even under a limited firing rate range.



Experimental Procedures

All animal procedures complied with US National Institutes of Health Guide for Care and Use of Laboratory Animals, and were approved by the Rockefeller University Institutional Animal Care and Use Committee (IACUC).

Animal Preparation.

Experiments were conducted on three male adult rhesus macaques, two of which were *Macaca mulatta* (monkey Q and M) and one of which was *Macaca fascicularis* (monkey J). All of them weighed 8-10 kg. Under general anesthesia and aseptic surgical conditions, the monkeys were implanted with Ultem headposts and recording chambers that were attached to the skull with dental acrylic and ceramic screws (Thomas Recording or Rogue Research). Following recovery, the monkeys were trained to maintain a fixation on a white spot (0.1-0.3 degrees) on a CRT (cathode ray tube, Iiyama, 36.6×27.4 cm; 1920×1440 pixels; 100 Hz refresh rate) screen for juice rewards. As the fixation deviated from tolerance window (3×3 degrees), the reward delivery stopped.

Monkey MRI.

Scanning was performed on a 3T MR scanner (TIM Trio with AC88 gradient insert; Siemens). Multi-echo sequence (EPI, TR 2 or 3 s, TE 30 ms, 64×64 matrix, 1.5 or 1.0 mm^3 voxels isotropic resolution) were acquired with field map to unwarp the image based on the B0-field inhomogeneities. Custom-made 8-channel surface coil was used, and MION contrast agent [186] was used to improve SNR. Face patches were defined by identifying regions responding significantly more to faces than non-face objects. Anatomical volumes at high spatial resolution (0.5 mm isotropic) with a T1-weighted inversion recovery sequence (MPRAGE) [187] were acquired using custom-made 1-channel coil to be overlayed with the functional data.

Face patch targeting-procedures.

We used either Caret [156] or Planner [158] software to overlay anatomical MRI and functional MRI in order to determine skull position and orientation of a recording chamber (Crist, [159]). A delrin-made recording grid (Crist, [159], 1 mm holes, 1 mm center-to-center distance) is placed inside the chamber to fix electrode insertion position. To compute which grid angle and grid hole to use within the recording chamber, we overlaid anatomical MRI acquired with a recording grid inside the recording chamber, filled with MR-visible silicone or gadolinium solution. The target face patch, MF, was accessible from only 1 or 2 holes (adjacent holes have 1 mm separation), because of blood vessels hampering trajectories.

Single-unit electrophysiological recordings and visual stimuli.

Extracellular recordings were conducted with 1-3 $M\Omega$ (measured at 1000 Hz) tungsten electrodes (FHC). The dura was penetrated using a metal guide tube

(23 gauge) through a grid hole [159]. The electrode was advanced using Narishige MO-97A drive until it reached the desired location. The electrophysiological signal was amplified and sorted online into single units using multiple discrimination windows (Blackrock Microsystems). These manually sorted clusters were compared with the clusters generated by offline clustering based on wavelet decomposition and superparamagnetic clustering (Waveclus algorithm [188].) for quality check. During the recordings, eye position of the animals was monitored by an infrared pupil tracking system (ETL-200, ISCAN Inc., Burlington, MA). The monkey was required to maintain fixation at the fixation spot throughout the experiments.

Identifying face cells.

As an electrode was advanced into MF, Face-Object-Body (FOB, 147 different stimuli containing 80 faces, 66 non-face objects and a gray square same color as the background. Average luminance was $\sim 18 \text{ cd/m}^2$). Stimuli were presented (for 200 ms without inter-stimulus intervals) to activate face neurons. The FOB stimuli were moved across the screen to estimate the position of the RF. Once a single neuron was isolated with its putative receptive field, the stimulus location was fixed to record neural responses to define the preferred face, the non-preferred face and the non-preferred object out of 147 FOB stimuli. The mean firing rate was calculated by custom made MATLAB script and the preferred face, non-preferred face and the non-preferred object were selected based on the mean firing rate. The stimulus size was 4×4 degrees of visual angles. The 147 FOB stimuli are not contrast-normalized: however, we compared neural responses to contrast-normalized FOB and un-normalized FOB, to see no noticeable difference in response patterns and magnitudes.

Receptive field mapping.

One of the defined preferred face, the non-preferred face or the non-preferred object was randomly presented at one location of hexagonal grid (18×24 degrees, 3 degrees of grid separation, see Fig. 2.8.) for 200 ms without inter-stimulus intervals. Stimulus size was magnified in proportion to the eccentricity (10 percent/degree) to compensate the cortical magnification factor¹ [189]. After the calculation, the center of the receptive field was selected manually to be used as the presentation location of the following experiments. The size of RF was estimated by calculating RF size index as follows: we calculated average firing rate across space within each of 3×3 square presentation grid placed at the center of the RF (each square took 4×4 degrees of visual angle). The calculated 9 average firing rate was sorted in descending order, then the first two (peak value) and the other seven (surround value) were averaged, respectively. RF size index was defined as the difference over sum of the two values:

$$I_{RF} = \frac{RF_{peak} - RF_{sur}}{RF_{peak} + RF_{sur}} \quad (A.1)$$

Acquiring neural responses to the multiple stimuli.

Upon the definition of the preferred/non-preferred stimuli, composite stimuli were created of 0, 1, 2, 4 or 8 non-preferred face(s), non-preferred-object(s) or preferred face(s) with or without the preferred face at the center (stimuli were generated by a custom MATLAB program after defining the RF center). When the preferred face was not presented at the center, a gray square (R128/G128/B128) was presented at the center which has the same color as

¹It was shown that the overall receptive field pattern was same regardless of the magnification [89].

the background screen. In the 8 non-preferred faces condition, we also used heterogeneous, conspecific face distractors. Namely, 8 least preferred faces were chosen from same species (monkey or human) as the selected preferred face. These heterogeneous, conspecific distractors were presented with 0, 2 or 4 degrees of separation from the center image. We also included a gray square in the stimulus set in order to calculate baseline responses. In total, 28 stimulus conditions were used. The stimuli were presented at 3×3 array, whose center was located at the defined receptive field center. In most of the experiments the preferred face was presented at the location defined as the receptive field center, but as a control we also presented the preferred face at surrounding locations (positional control). Each stimulus was 4×4 degrees (thus, the 3×3 stimulus array spans from 12×12 degrees to 20×20 degrees) and presented for 400 ms with 200 ms inter-stimulus gap. In 1 distractor conditions, the distractor was placed randomly at one of 8 possible presentation locations. In 2 and 4 distractors conditions, the distractors were placed randomly at symmetric positions of 8 possible presentation locations. Note that because of the variable sizes of MF neurons' receptive fields, some of the stimuli could be presented outside of the receptive field. All stimuli were presented on a CRT screen placed 57 cm in front of the monkey (such that 1 cm on the screen approximately equal to 1 degree of visual angle. see Appendix E). All stimuli were controlled by custom software written in C (Visiko) running on a windows PC. Stimulus presentation timestamps recorded by Visiko refers to the start of the display frame during which the image was presented. Thus, given the 100 Hz refresh rate of the CRT screen, the actual stimulus presentation could have up to 10 ms ($= 1 \text{ second} / 100 \text{ Hz}$) delay depending on the stimulus position.

Data analysis.

Recording data were analyzed with custom programs written in MATLAB. Out of 114 recorded neurons, neurons with average responses to their preferred faces did not exceed 3 standard deviations above baseline response were omitted from the rest of analyses. The baseline responses were calculated as the mean firing rate within 100-499 ms after the onset of gray background presentation.

FOB stimuli.

The time window in which mean firing rate was calculated was defined manually according to the neuron's response latency. In Fig. 1, the mean firing rate was calculated in the window of 100-299ms after the stimulus onset.

Response latency calculation.

Response onset was defined as the time when the response to the neuron's preferred face exceeded 3 standard deviations above baseline response. The baseline firing rate was calculated during presentation of a gray square that has the same color as screen background.

Definition of the early and the late phase. The early/late phase encompasses the first peak and the late plateau of typical MF neuron's response, respectively. To equalize noises within the time window as much as possible, we defined the early/late phase such that the number of spikes within each phase became approximately equal. We used the population average response to the preferred face to calculate a response latency, and with the latency the Early and the Late phase were defined as 0-99 ms and 150-399 ms after the response onset, respectively.

Receptive fieldmapping.

The receptive field was drawn by interpolating the mean firing rate (calculated using the same time window as FOB) at each presentation point, by custom MATLAB code.

Regression analysis.

To fit data from Cat/Num experiment, we used the equation,

$$R = \frac{1}{1 + \alpha x} \quad (3.1)$$

Where x is the number of distractors, and α is a free parameter. The mean responses were divided by the response to the preferred face, such that all the response could be compared relative to the response to the preferred face. In this equation, the integration of multiple stimuli were approximately average when $\alpha=1$, and were approximately maximum when $\alpha=0$, assuming that the responses to the non-preferred images were close to zero. Regression was conducted by MATLAB function `nlinfit`.

Divisive normalization model.

See Appendix B.

B

Divisive normalization model

We designed the model equation such that the output from a single neuron is the sum of direct drives elicited by each visual stimulus, where each drive is weighted by the population average of the drives. The weights are normalized such that the sum of them equals 1, thus the equation works as a weighted average based on population activity. The response of neuron i is therefore

$$s_i^{\text{pref+dist}} = \frac{\sum_{p=1}^l s_i^{\text{pref}}(x_p) p_i^{\text{dist}}(x_p) + \sum_{q=1}^m s_i^{\text{dist}}(y_q) p_i^{\text{dist}}(y_q)}{\sum_{p=1}^l p_i^{\text{dist}}(x_p) + \sum_{q=1}^m p_i^{\text{dist}}(y_q)} \quad (\text{B.1})$$

where

$$\begin{aligned} p_i^{\text{pref}}(x_p) &= \frac{1}{N} \sum_{i=1}^N s_i^{\text{pref}}(x_p) \\ p_i^{\text{dist}}(y_q) &= \frac{1}{N} \sum_{i=1}^N s_i^{\text{dist}}(y_q) \end{aligned} \quad (\text{B.2})$$

Here $s_i^{\text{pref}}(x_p)$ and $s_i^{\text{dist}}(y_q)$ describe the direct drives to neuron i , evoked by the preferred face and the distractor, respectively, as a function of presentation location. Variables x_p and y_q stand for the presentation location of preferred face and distractor, respectively, and l and m are the total number of preferred faces and distractors presented. N is the total number of recorded neurons.

Since experimentally observed variables are in principle the result of this normalization, we do not have direct access to internal variables $s_i^{\text{pref}}(x_p)$ and $s_i^{\text{dist}}(y_q)$. To find these values, we consider the RF mapping experiment where non-combined stimuli were used. In RF experiment only one preferred face, one non-preferred face or one non-preferred object is presented. For the case of only one preferred face being used, the second term of both denominator and numerator of Equation B.1 are zero, therefore we get

$$\begin{aligned} S_i^{\text{pref}}(x_p) &= \frac{s_i^{\text{pref}}(x_p)p_i^{\text{pref}}(x_p)}{p_i^{\text{pref}}(x_p)} \\ &= s_i^{\text{pref}}(x_p) \end{aligned} \tag{B.3}$$

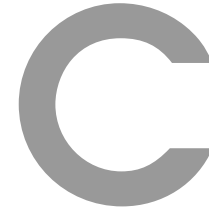
where $S_i^{\text{pref}}(x_p)$ is the value of RF of neuron i for its preferred face at location x_p . The same logic can be applied to the case when a single distractor is presented. Thus, the internal variables $s_i^{\text{pref}}(x_p)$ and $s_i^{\text{dist}}(y_q)$ can be substituted by the measured RF values, $S_i^{\text{pref}}(x_p)$ and $S_i^{\text{dist}}(y_q)$, respectively. Using Equation B.2, $p_i^{\text{pref}}(x_p)$ and $p_i^{\text{dist}}(y_q)$ are similarly substituted by the calculated population activities, $P_i^{\text{pref}}(x_p)$ and $P_i^{\text{dist}}(y_q)$. The resulting equation of the pure-prediction model is therefore

$$S_i^{\text{pref+dist}} = \frac{\sum_{p=1}^l S_i^{\text{pref}}(x_p) P_i^{\text{dist}}(x_p) + \sum_{q=1}^m S_i^{\text{dist}}(y_q) P_i^{\text{dist}}(y_q)}{\sum_{p=1}^l P_i^{\text{dist}}(x_p) + \sum_{q=1}^m P_i^{\text{dist}}(y_q)} \quad (5.1)$$

However to calculate for example P_i^{pref} , we need to know the RFs of all the recorded neurons for the preferred face of the i -th neuron. But stimulus preferences are different from neuron to neuron, and we mapped only 3 RFs per each neuron (RF for the neuron's preferred face, RF for its non-preferred face and RF for non-preferred object). Many RFs needed to calculate the population average are thus missing. We go about this problem by assuming that RF size and shape does not change with stimulus identity. This means we can obtain the RF of a given neuron for any of the FOB stimuli by using the preferred face RF of each neuron and scaling the magnitude by the neuron's normalized mean response to FOB stimuli.

To further improve the predictive power of Equation 5.1, we introduce a free parameter, exponent n . The exponent can change neuronal computation smoothly from weighted average of all stimuli to winner-take-all of one of the stimuli. The following Equation 5.2 is used as the post-diction model.

$$S_i^{\text{pref+dist}} = \frac{\sum_{p=1}^l S_i^{\text{pref}}(x_p) \{P_i^{\text{dist}}(x_p)\}^n + \sum_{q=1}^m S_i^{\text{dist}}(y_q) \{P_i^{\text{dist}}(y_q)\}^n}{\sum_{p=1}^l \{P_i^{\text{pref}}(x_p)\}^n + \sum_{q=1}^m \{P_i^{\text{dist}}(y_q)\}^n} \quad (5.2)$$



Converting anatomical plane coordinate to stereotaxic arm coordinate

We calculate the angle of recording chamber with planning software such as Caret [156], OsiriX [157] or Planner [158] to define desired angles within two of coronal, horizontal or sagittal planes (Fig. C.1A. Angle in the third plane is complement and automatically specified) . However, common stereotaxic arm does not allow users to make a defined angle within each anatomical plane. Rather, it has only two degrees of freedom: tilting the stereotaxic arm and rotating the base of the stereotaxic arm (Fig. C.1B). Thus, in order to implant the planned recording chamber angle using a stereotaxic arm, coordinates in the anatomical sections have to be converted into stereotaxic arm coordinate such that tilting and rotation of a stereotaxic arm ends up making desired angles within anatomical planes as planned.

In Fig. C.1C, the line AB and AC indicate the direction of stereotaxic arm base and stereotaxic arm, respectively. Within the planning software, the

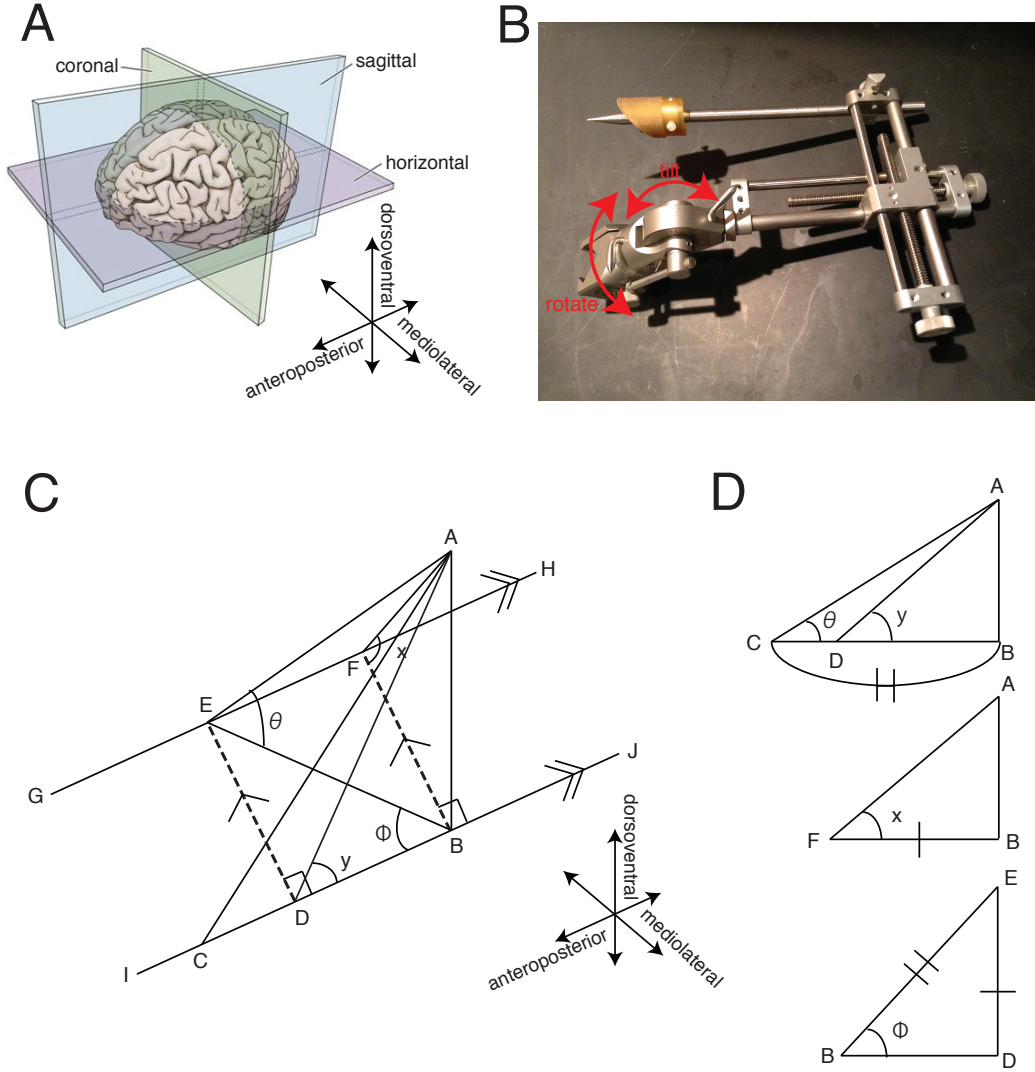


Figure C.1: Converting anatomical planes coordinate into stereotaxic coordinates. (A) Definition of the three anatomical sections: coronal, horizontal and sagittal section. Note that shown is a human brain, not a macaque brain. Adapted from [190]. (B) Kopf Stereotaxic arm (David Kopf Instruments) with a recording chamber loaded onto a pointer. The arm has two degree of freedoms, (1) rotation of the base, and (2) tilting the arm. (C) Schematic diagram to calculate rotation and tilting angle in order to achieve desired angle within anatomical planes. The line AB and AC indicates the direction of the stereotaxic arm base and stereotaxic arm, respectively. θ and ϕ are tilting and rotation angle on the stereotaxic arm (ϕ is defined from anteroposterior axis as shown). x and y are desired angle within a coronal and sagittal plane, respectively. (D) Relevant triangles required for calculating θ and ϕ , excerpted from (C). Vertical bars | and || indicates pairs of sides with equal length.

chamber angles x and y are defined within coronal and sagittal plane, and θ and ϕ are tilting and rotation angle on the stereotaxic arm and base needed in order to achieve the desired angle within anatomical planes. Fig. C.1D shows three excerpted triangles needed for calculating θ and ϕ . Our goal is to express θ and ϕ in terms of x and y . Fig. C.1D provides the following relationships:

$$\begin{aligned}
 \tan y &= \frac{AB}{DB} \\
 \tan \theta &= \frac{AB}{CB} \\
 &= \frac{AB}{EB} \\
 \tan x &= \frac{AB}{FB} \\
 &= \frac{AB}{ED} \\
 \tan \phi &= \frac{ED}{BD} \\
 \sin \phi &= \frac{ED}{EB}
 \end{aligned} \tag{C.1}$$

From the relationships in Equations C.1, θ and ϕ are expressed in terms of x and y as follows:

$$\begin{aligned}
\tan x \tan \phi &= \frac{AB}{ED} \frac{ED}{BD} \\
&= \frac{AB}{BD} \\
&= \tan y \\
\tan \phi &= \frac{\tan y}{\tan x} \\
\phi &= \arctan \left(\frac{\tan y}{\tan x} \right)
\end{aligned} \tag{C.2}$$

$$\begin{aligned}
\frac{\tan \theta}{\tan x} &= \frac{AB}{EB} \frac{ED}{AB} \\
&= \frac{ED}{EB} \\
&= \sin \theta \\
\tan \theta &= \tan x \sin \phi
\end{aligned} \tag{C.3}$$

$$\begin{aligned}
\theta &= \arctan (\tan x \sin \phi) \\
&= \arctan \left(\tan x \sin \left(\arctan \left(\frac{\tan y}{\tan x} \right) \right) \right)
\end{aligned}$$



Converting pixels to degrees of visual angle

To plot RF in unit of degrees, pixel size on the stimulus presentation screen must be converted into unit of degrees.

In the Fig. D.1, the screen-to-eye distance s , image viewing angle y and image pixel size x is defined, in addition to these parameters, the diagonal length of the screen is defined in pixels (d) and in cm (c). d and c can be written as:

$$d = \sqrt{w^2 + h^2}$$

$$c = \sqrt{a^2 + b^2}$$

where w/h and a/b are screen width/height in pixels and in cm, respectively. Using d and c , image size on the screen can be converted from pixels to cm. Assuming the subject is looking at the presentation screen perpendicularly (this assumption is valid because the subject is required to fixate on the

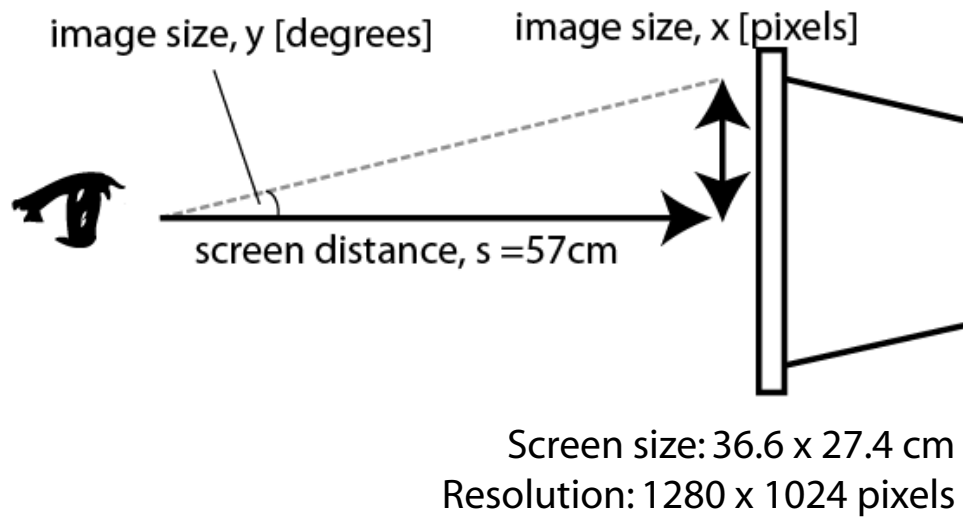


Figure D.1: Converting pixels to degrees. Screen-to-eye distance (s) is fixed to 57 cm. Image size on the screen is x pixels, and subject's viewing angle of the image is y degrees. Screen width is 36.6 cm (w) corresponds to 1280 pixels (a), and height is 27.4 cm (h) corresponds to 1024 pixels (b). See the text how to describe the x in terms of y .

center dot of the screen throughout the experiment), tangent function is used to relate the viewing angle y and the image pixel size x :

$$\tan y = \frac{x \frac{c}{d}}{s} \quad (\text{D.1})$$



Finding the screen distance equalizing visual angle and image size

The viewing distance (screen-to-eye distance) was fixed to 57 cm such that 1 cm on the stimulus presentation screen approximately equals to 1 degree of animal's viewing angle at small viewing angle. Fig. E.1 shows the schematic illustration of the problem. We need to find a screen distance s , which ensures $x = y$, where x and y are image size on the screen (cm) and image viewing angle, y (degrees), respectively.

Equation E.1 shows the calculation of s . Note that approximations $\sin \theta \approx \theta$ and $\cos \theta \approx 1$ were used given that the image viewing angle is sufficiently small.

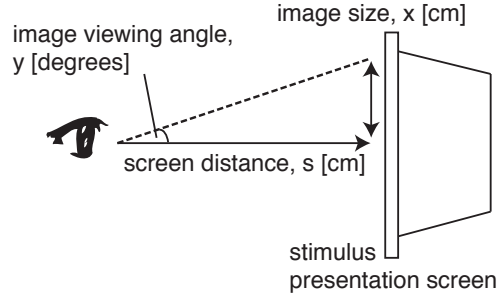


Figure E.1: Relating viewing distance to image size on the screen. The distance from the subject's eye to stimulus presentation monitor (screen distance s cm) affects the relation between the image viewing angle (y degree) and the image size (x cm) on the screen. See Equation E.1 to find the screen distance approximately equate x and y .

$$\begin{aligned}
 \tan\left(y \frac{\pi}{180}\right) &= \frac{x}{s} \\
 s &= \frac{x}{\tan\left(y \frac{\pi}{180}\right)} \\
 &\approx \frac{x}{x \frac{\pi}{180}} \\
 &= 57.30
 \end{aligned}
 \tag{E.1}$$

F

Population average without latency subtraction

111recorded MF neurons were averaged across trials and neurons to plot population average firing rate (Fig. F.1). For details and statistics, see Chapter 3 and Appendix J.

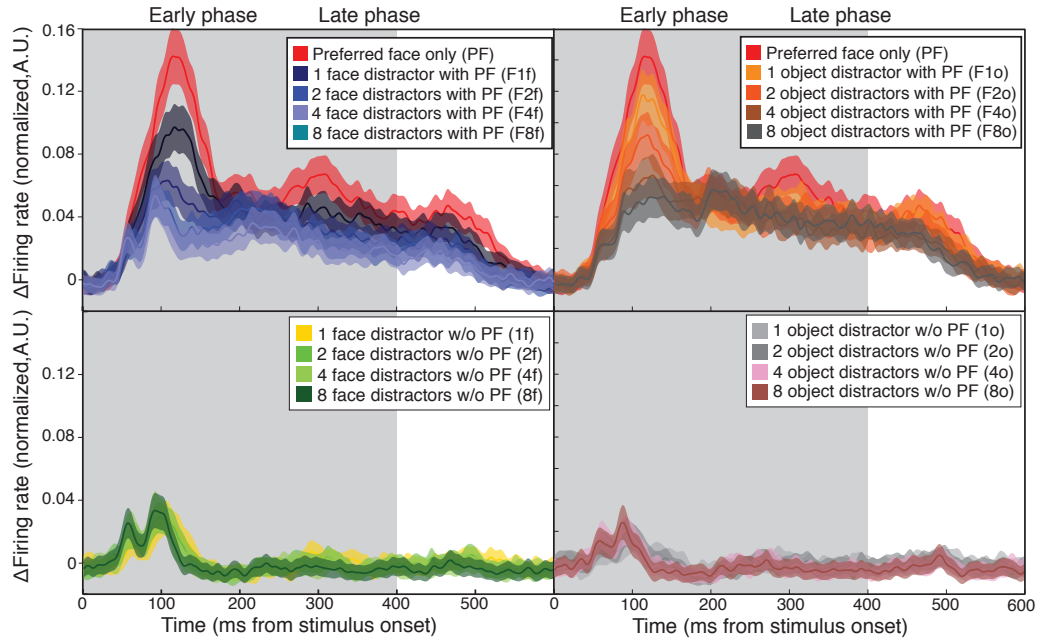


Figure F.1: Population average without latency subtraction. Mean firing rates of 111 MF neurons were averaged across neurons without correcting latency difference. Gray shadow indicates the stimulus presentation duration.



Population average firing rate to multiple preferred faces.

Responses to increasing number of preferred faces were significantly decreased in MF neurons, but not significantly reduced in AL neurons. Interpretation in terms of divisive normalization is that MF neurons have smaller RF than AL neurons, entailing higher chance of having preferred faces outside of the RF resulting in (surround) suppression. Fig. G.1 shows the response latency-subtracted population averages. One-way ANOVA was used to test whether increasing number of preferred faces change the mean firing rate. For statistics, see the Fig. G.1 legend.

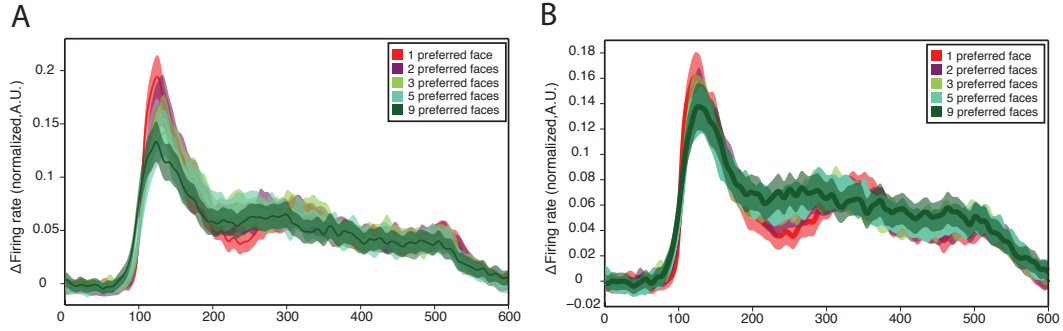


Figure G.1: Population average firing rate to multiple (1, 2, 3, 5 and 9) preferred faces. Response latencies were subtracted and averaged across neurons. (A) MF neurons (1,9 preferred faces: $N=111$, 2,3,5 preferred faces: $N=68$). Responses were significantly reduced in the early phase (one-way ANOVA, $p = 0.0173 < 0.05$) by increasing number of preferred face. No significant difference in the late phase ($p = 0.664 > 0.05$). (B) AL neurons ($N=109$ from two monkeys). Responses were not significantly reduced by increasing number of preferred faces either in the early (one-way ANOVA, $p = 0.712 > 0.05$) and the late ($p = 0.594 > 0.05$) phase.



Visualizing the population average in 2D space

One way to visualize the population average activity in two-dimensional space is to project each data point to the line of unity.

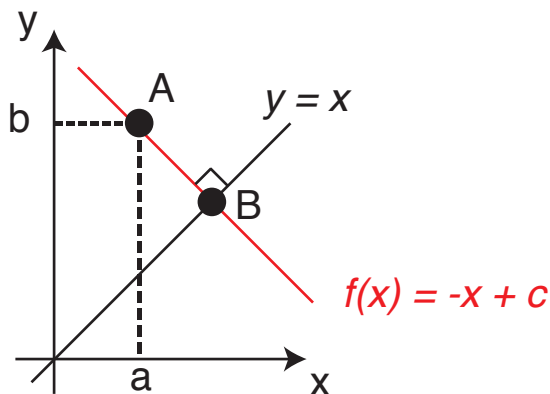


Figure H.1: Projection to diagonal. The hypothetical data point A represents the response of a neuron x ($x = a$) and a neuron y ($y = b$). B is the data point A projected to the diagonal line of unity. The red line $f(x)$ can be written as $f(x) = -x + c$ where c is a constant. The red line crosses both A and B, and meets perpendicular to the line of unity.

In Fig. H.1, an example data point A has values $(x, y) = (a, b)$, representing activities of two hypothetical neurons. Thus, the average response of the two neurons is $\frac{a+b}{2}$. To project the point A to the line of unity, we solve simultaneous equation consists of the line of unity and perpendicular line crossing the point A ($f(x)$ in Fig. H.1). The unknown intercept c can be find given that $f(x)$ crosses the point A:

$$\begin{aligned} f(x) &= -x + c \\ b &= -a + c \\ c &= a + b \end{aligned} \tag{H.1}$$

Then we solve the simultaneous equation to find the point B. Both x and y values of B is $\frac{a+b}{2}$, representing a population average.

$$\begin{cases} y = x \\ y = -x + (a + b) \end{cases} \tag{H.2}$$

$$\begin{aligned} x &= \frac{a + b}{2} \\ y &= \frac{a + b}{2} \end{aligned}$$



Relationship between PCA and SVD

PCA and SVD are two closely related methods. In fact, PCA is often computed directly from SVD [191, 192, 193, 194] as follows.

We first start with applying PCA to a centered (mean set to zero) $n \times c$ data matrix X . PCA calculates covariance matrix of the data matrix X and diagonalize the covariance matrix with eigenvectors. The covariance matrix is calculated as follows:

$$\begin{aligned} Cov &= E(XX^T) - E(X)E(X)^T \\ &= \frac{1}{n}XX^T \\ &=\propto XX^T \end{aligned}$$

Where n is number of data points and $E(X) = 0$ because X is centered. Similarly, covariance matrix of transpose of X ¹ is proportional to $X^T X$. Since

¹By taking the transposition of X , the dimension to be reduced can be changed. XX^T and $X^T X$ yields matrices of $n \times n$ and $c \times c$, respectively, resulting in reducing 1st and 2nd

the covariance matrix is symmetric, the matrix is diagonalizable by matrices C, D which consists of eigenvectors of X . PCA decomposes the covariance matrix into the product of three matrices:

$$\begin{aligned} XX^T &= ACA^T \\ X^T X &= BDB^T \end{aligned} \tag{I.1}$$

Where C, D is a diagonal matrix whose diagonal elements are eigenvalues. SVD on the other hand, decompose the original data matrix X instead of covariance matrix:

$$X = USV^T$$

Where U and V are $n \times n$ and $c \times c$ orthonormal matrices, respectively. Diagonal matrix S has the same size $n \times c$ as the data matrix X . The diagonal element of S is singular values, whose square are equal to the eigenvalues:

$$\begin{aligned} XX^T &= USV^T(USV^T)^T \\ &= USV^T V S^T U^T \\ &= US^2 U^T \\ X^T X &= (USV^T)^T USV^T \\ &= (VSU^T USV^T \\ &= VS^2 V^T \end{aligned} \tag{I.2}$$

dimension of original data matrix, respectively.

Thus, Equation I.1 and I.2 shows that depending on calculating covariance matrix of either X or X^T (in other words, depending on reducing whether 1st or 2nd dimension), either the matrix U or V is equal to eigenvectors, and square of the singular values are equal to eigenvalues.



Tukey's Post-hoc test results

Table J.1: MF single cell example of Cat/Num experiment, face distractor conditions at the early phase

PF+1FD	<div><div>*</div><div>0.0326</div></div>			
PF+2FD	<div><div>***</div><div>$2.25 \cdot 10^{-7}$</div></div>	<div><div></div><div>0.0512</div></div>		
PF+4FD	<div><div>***</div><div>$9.92 \cdot 10^{-9}$</div></div>	<div><div>***</div><div>$1.79 \cdot 10^{-5}$</div></div>	<div><div></div><div>0.278</div></div>	
PF+8FD	<div><div>***</div><div>$9.92 \cdot 10^{-9}$</div></div>	<div><div>***</div><div>$9.92 \cdot 10^{-9}$</div></div>	<div><div>***</div><div>$1.22 \cdot 10^{-6}$</div></div>	<div><div>**</div><div>$5.70 \cdot 10^{-3}$</div></div>
	PF only	PF+1FD	PF+2FD	PF+4FD

Table J.2: MF single cell example of Cat/Num experiment, face distractor conditions at the late phase

PF+1FD	** $9.09 \cdot 10^{-3}$			
PF+2FD	** $2.52 \cdot 10^{-3}$	0.997		
PF+4FD	*** $2,34 \cdot 10^{-8}$	0.0570	0.141	
PF+8FD	*** $9.92 \cdot 10^{-9}$	*** $3.20 \cdot 10^{-7}$	*** $2.38 \cdot 10^{-6}$	0.0270
	PF only	PF+1FD	PF+2FD	PF+4FD

Table J.3: MF single cell example of Cat/Num experiment, object distractor conditions at the early phase

PF+1OD	0.243			
PF+2OD	*** $8.82 \cdot 10^{-4}$	0.339		
PF+4OD	*** $5.61 \cdot 10^{-8}$	** $1.29 \cdot 10^{-3}$	0.289	
PF+8OD	*** $9.92 \cdot 10^{-9}$	*** $6.55 \cdot 10^{-7}$	** $3.59 \cdot 10^{-3}$	0.504
	PF only	PF+1OD	PF+2OD	PF+4OD

Table J.4: MF single cell example of Cat/Num experiment, object distractor conditions at the late phase

PF+1OD	0.639			
PF+2OD	0.315	0.986		
PF+4OD	0.103	0.831	0.983	
PF+8OD	0.0951	0.821	0.981	0.999
	PF only	PF+1OD	PF+2OD	PF+4OD

Table J.5: MF population average of Cat/Num experiment, face distractor conditions at the early phase

PF+1FD	*** $2.89 \cdot 10^{-7}$			
PF+2FD	*** $9.92 \cdot 10^{-9}$	*** $1.90 \cdot 10^{-5}$		
PF+4FD	*** $9.92 \cdot 10^{-9}$	*** $1.03 \cdot 10^{-8}$	0.132	
PF+8FD	*** $9.92 \cdot 10^{-9}$	*** $9.92 \cdot 10^{-9}$	** $2.26 \cdot 10^{-3}$	0.673
	PF only	PF+1FD	PF+2FD	PF+4FD

Table J.6: MF population average of Cat/Num experiment, face distractor conditions at the late phase

PF+1FD	*			
	0.0120			
PF+2FD	***			
	$7.89 \cdot 10^{-7}$	0.190		
PF+4FD	***	*		
	$1.11 \cdot 10^{-8}$	0.0107	0.825	
PF+8FD	***	***		
	$9.92 \cdot 10^{-9}$	$3.95 \cdot 10^{-5}$	0.105	0.643
	PF only	PF+1FD	PF+2FD	PF+4FD

Table J.7: MF population average of Cat/Num experiment, object distractor conditions at the early phase

PF+1OD	**			
	$9.56 \cdot 10^{-3}$			
PF+2OD	***			
	$1.35 \cdot 10^{-7}$	0.109		
PF+4OD	***	***		
	$9.92 \cdot 10^{-9}$	$9.43 \cdot 10^{-6}$	0.0956	
PF+8OD	***	***	**	
	$9.92 \cdot 10^{-9}$	$1.37 \cdot 10^{-8}$	$1.19 \cdot 10^{-3}$	0.656
	PF only	PF+1OD	PF+2OD	PF+4OD

Table J.8: MF population average of Cat/Num experiment, object distractor conditions at the late phase

PF+1OD	0.185			
PF+2OD	* 0.0161	0.887		
PF+4OD	** $1.37 \cdot 10^{-3}$	0.488	0.959	
PF+8OD	** $9.49 \cdot 10^{-3}$	0.813	0.999	0.985
	PF only	PF+1OD	PF+2OD	PF+4OD

Table J.9: MF single cell example of Prox/Ecc experiment, RF center at the early phase

PF+LD	*** $3.77 \cdot 10^{-9}$		
PF+MD	*** $3.77 \cdot 10^{-9}$	0.823	
PF+SD	*** $3.77 \cdot 10^{-9}$	*** $4.92 \cdot 10^{-4}$	0.0111
	PF only	PF+LD	PF+MD

Table J.10: MF single cell example of Prox/Ecc experiment, RF center at the late phase

PF+LD	0.901		
PF+MD	0.415	0.639	
PF+SD	0.537	0.802	0.993
	PF only	PF+LD	PF+MD

Table J.11: MF single cell example of Prox/Ecc experiment, RF periphery at the early phase

PF+LD	*** $2.37 \cdot 10^{-4}$		
PF+MD	*** $3.64 \cdot 10^{-4}$	0.997	
PF+SD	*** $3.26 \cdot 10^{-8}$	0.0782	* 0.0401
	PF only	PF+LD	PF+MD

Table J.12: MF single cell example of Prox/Ecc experiment, RF periphery at the late phase

PF+LD	*** $5.37 \cdot 10^{-7}$		
PF+MD	*** $4.06 \cdot 10^{-9}$	0.364	
PF+SD	*** $8.22 \cdot 10^{-8}$	0.974	0.621
	PF only	PF+LD	PF+MD

Table J.13: MF population average of Prox/Ecc experiment, RF center at the early phase

PF+LD	*** $9.78 \cdot 10^{-9}$		
PF+MD	*** $3.77 \cdot 10^{-9}$	0.441	
PF+SD	*** $3.77 \cdot 10^{-9}$	*** $2,26 \cdot 10^{-4}$	* 0.0435
	PF only	PF+LD	PF+MD

Table J.14: MF population average of Prox/Ecc experiment, RF center at the late phase

PF+LD	*** $1.79 \cdot 10^{-6}$		
PF+MD	*** $3.87 \cdot 10^{-9}$	0.376	
PF+SD	*** $3.77 \cdot 10^{-9}$	** $1.65 \cdot 10^{-3}$	0.180
	PF only	PF+LD	PF+MD

Table J.15: MF population average of Prox/Ecc experiment, RF periphery at the early phase

PF+LD	*** $6.86 \cdot 10^{-9}$		
PF+MD	*** $3.99 \cdot 10^{-9}$	0.978	
PF+SD	*** $6.01 \cdot 10^{-9}$	0.999	0.985
	PF only	PF+LD	PF+MD

Table J.16: MF population average of Prox/Ecc experiment, RF periphery at the late phase

PF+LD	*** $4.25 \cdot 10^{-9}$		
PF+MD	*** $3.77 \cdot 10^{-9}$	0.876	
PF+SD	*** $3.78 \cdot 10^{-9}$	0.960	0.994
	PF only	PF+LD	PF+MD

Table J.17: MF population average of Prox/Ecc experiment, distractor only control, RF center at the early phase

MD only	0.989	
SD only	* 0.0414	0.0590
	LD only	MD only

Table J.18: MF population average of Prox/Ecc experiment, distractor only control, RF center at the early phase

MD only	0.981	
SD only	0.441	0.555
	LD only	MD only

Table J.19: MF population average of Prox/Ecc experiment, distractor only control, RF periphery at the early phase

MD only	0.354	
SD only	<div> <div>**</div> <div>$3.69 \cdot 10^{-3}$</div> </div>	0.156
	LD only	MD only

Table J.20: MF population average of Prox/Ecc experiment, distractor only control, RF periphery at the early phase

MD only	0.442	
SD only	0.0704	0.5860
	LD only	MD only

Table J.21: AL population average of Cat/Num experiment, face distractor conditions at the early phase

PF+1FD	*** $1.14 \cdot 10^{-4}$			
PF+2FD	*** $9.92 \cdot 10^{-9}$	** $5.35 \cdot 10^{-3}$		
PF+4FD	*** $9.92 \cdot 10^{-9}$	*** $1.61 \cdot 10^{-8}$	* 0.0476	
PF+8FD	*** $9.92 \cdot 10^{-9}$	*** $9.92 \cdot 10^{-9}$	*** $2.60 \cdot 10^{-4}$	0.593
	PF only	PF+1FD	PF+2FD	PF+4FD

Table J.22: AL population average of Cat/Num experiment, face distractor conditions at the late phase

PF+1FD	0.312			
PF+2FD	** $6.98 \cdot 10^{-3}$	0.594		
PF+4FD	*** $1.57 \cdot 10^{-4}$	0.114	0.876	
PF+8FD	*** $2.49 \cdot 10^{-6}$	* 0.0102	0.375	0.916
	PF only	PF+1FD	PF+2FD	PF+4FD

Table J.23: AL population average of Cat/Num experiment, object distractor conditions at the early phase

PF+1OD	0.254			
PF+2OD	*** $2.53 \cdot 10^{-4}$	0.187		
PF+4OD	*** $7.43 \cdot 10^{-7}$	** $7.07 \cdot 10^{-3}$	0.765	
PF+8OD	*** $1.01 \cdot 10^{-8}$	*** $2.73 \cdot 10^{-5}$	0.0889	0.672
	PF only	PF+1OD	PF+2OD	PF+4OD

Table J.24: AL population average of Cat/Num experiment, object distractor conditions at the late phase

PF+1OD	0.506			
PF+2OD	0.194	0.9786		
PF+4OD	* 0.0179	0.569	0.891	
PF+8OD	* 0.0144	0.524	0.863	0.999
	PF only	PF+1OD	PF+2OD	PF+4OD

Table J.25: AL population average of Prox/Ecc experiment, RF center at the early phase

PF+LD	*** $2.29 \cdot 10^{-5}$		
PF+MD	*** $8.40 \cdot 10^{-9}$	0.420	
PF+SD	*** $3.77 \cdot 10^{-9}$	*** $3.42 \cdot 10^{-5}$	* 0.0141
	PF only	PF+LD	PF+MD

Table J.26: AL population average of Prox/Ecc experiment, RF center at the late phase

PF+LD	*** $2.10 \cdot 10^{-5}$		
PF+MD	*** $7.62 \cdot 10^{-8}$	0.713	
PF+SD	*** $3.77 \cdot 10^{-9}$	0.063	0.488
	PF only	PF+LD	PF+MD

Table J.27: AL population average of Prox/Ecc experiment, RF periphery at the early phase

PF+LD	*** $1.56 \cdot 10^{-7}$		
PF+MD	*** $1.78 \cdot 10^{-6}$	0.971	
PF+SD	*** $4.03 \cdot 10^{-7}$	0.998	0.993
	PF only	PF+LD	PF+MD

Table J.28: AL population average of Prox/Ecc experiment, RF periphery at the late phase

PF+LD	*** $9.06 \cdot 10^{-7}$		
PF+MD	*** $7.86 \cdot 10^{-9}$	0.796	
PF+SD	*** $3.90 \cdot 10^{-9}$	0.479	0.955
	PF only	PF+LD	PF+MD

Table J.29: AL population average of Prox/Ecc experiment, distractor only control, RF center at the early phase

MD only	0.468	
SD only	* 0.0153	0.246
	LD only	MD only

Table J.30: AL population average of Prox/Ecc experiment, distractor only control, RF center at the early phase

MD only	0.924	
SD only	0.139	0.281
	LD only	MD only

Table J.31: AL population average of Prox/Ecc experiment, distractor only control, RF periphery at the early phase

MD only	0.583	
SD only	0.846	0.899
	LD only	MD only

Table J.32: AL population average of Prox/Ecc experiment, distractor only control, RF periphery at the early phase

MD only	0.797	
SD only	0.863	0.992
	LD only	MD only

Table J.33: MF population average without latency subtraction, Cat/Num experiment, face distractor conditions at the early phase

PF+1FD	*** $3.15 \cdot 10^{-4}$			
PF+2FD	*** $9.92 \cdot 10^{-9}$	** $2.02 \cdot 10^{-3}$		
PF+4FD	*** $9.92 \cdot 10^{-9}$	*** $6.95 \cdot 10^{-8}$	0.213	
PF+8FD	*** $9.92 \cdot 10^{-9}$	*** $9.92 \cdot 10^{-9}$	** $2.85 \cdot 10^{-3}$	0.568
	PF only	PF+1FD	PF+2FD	PF+4FD

Table J.34: MF population average without latency subtraction, Cat/Num experiment, face distractor conditions at the late phase

PF+1FD	** $1.77 \cdot 10^{-3}$			
PF+2FD	*** $1.67 \cdot 10^{-8}$	0.106		
PF+4FD	*** $9.92 \cdot 10^{-9}$	** $1.29 \cdot 10^{-3}$	0.641	
PF+8FD	*** $9.92 \cdot 10^{-9}$	*** $5.10 \cdot 10^{-6}$	0.073	0.747
	PF only	PF+1FD	PF+2FD	PF+4FD

Table J.35: MF population average without latency subtraction, Cat/Num experiment, object distractor conditions at the early phase

PF+1OD	0.207			
PF+2OD	*** $1.74 \cdot 10^{-4}$	0.196		
PF+4OD	*** $1.88 \cdot 10^{-8}$	*** $6.24 \cdot 10^{-4}$	0.354	
PF+8OD	*** $9.92 \cdot 10^{-9}$	*** $1.74 \cdot 10^{-6}$	* 0.0183	0.733
	PF only	PF+1OD	PF+2OD	PF+4OD

Table J.36: MF population average without latency subtraction, Cat/Num experiment, object distractor conditions at the late phase

PF+1OD	0.157			
PF+2OD	$02.94 \cdot 10^{-3}$	0.667		
PF+4OD	*** $6.95 \cdot 10^{-5}$	0.169	0.900	
PF+8OD	*** $6.61 \cdot 10^{49}$	0.423	0.996	0.986
	PF only	PF+1OD	PF+2OD	PF+4OD

Bibliography

- [1] G. Yovel and W. A. Freiwald, “Face recognition systems in monkey and human: are they the same thing?,” *F1000Prime Rep* **5** (2013) 10.
- [2] T. Allison, A. Puce, and G. McCarthy, “Social perception from visual cues: role of the STS region,” *Trends Cogn. Sci. (Regul. Ed.)* **4** no. 7, (Jul, 2000) 267–278.
- [3] A. R. Damasio, H. Damasio, and G. W. Van Hoesen, “Prosopagnosia: anatomic basis and behavioral mechanisms,” *Neurology* **32** no. 4, (Apr, 1982) 331–341.
- [4] M. Sachse, S. Schlitt, D. Hainz, A. Ciaramidaro, H. Walter, F. Poustka, S. Bolte, and C. M. Freitag, “Facial emotion recognition in paranoid schizophrenia and autism spectrum disorder,” *Schizophr. Res.* **159** no. 2-3, (Nov, 2014) 509–514.
- [5] Z. Zhu, P. Luo, X. Wang, and X. Tang, “Deep learning identity-preserving face space,” in *The IEEE International Conference on Computer Vision (ICCV)*. December, 2013.
- [6] R. Jenkins, D. White, X. Van Montfort, and A. Mike Burton, “Variability in photos of the same face,” *Cognition* **121** no. 3, (Dec., 2011) 313–323. <http://linkinghub.elsevier.com/retrieve/pii/S0010027711002022>.
- [7] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June, 2014.

- [8] X. Li, G. Mori, and H. Zhang, "Expression-invariant face recognition with expression classification," in *Computer and Robot Vision, 2006. The 3rd Canadian Conference on*, pp. 77–77. June, 2006.
- [9] Z. Li, U. Park, and A. Jain, "A discriminative model for age invariant face recognition," *Information Forensics and Security, IEEE Transactions on* **6** no. 3, (Sept, 2011) 1028–1037.
- [10] P. Sinha and T. Poggio, "'United' we stand," *Perception* **31** no. 1, (2002) 133.
- [11] J. I. Flombaum and L. R. Santos, "Rhesus monkeys attribute perceptions to others," *Curr. Biol.* **15** no. 5, (Mar, 2005) 447–452.
- [12] A. A. Ghazanfar and L. R. Santos, "Primate brains in the wild: the sensory bases for social interactions," *Nat. Rev. Neurosci.* **5** no. 8, (Aug, 2004) 603–616.
- [13] L. A. Parr and M. Heintz, "Facial expression recognition in rhesus monkeys, *Macaca mulatta*," *Anim Behav* **77** no. 6, (Jun, 2009) 1507–1513.
- [14] S. Kanazawa, "Recognition of facial expressions in a japanese monkey (*macaca fuscata*) and humans (*homo sapiens*)," *Primates* **37** no. 1, (1996) 25–38. <http://dx.doi.org/10.1007/BF02382917>.
- [15] V. Corkum and C. Moore, "The origins of joint visual attention in infants," *Dev Psychol* **34** no. 1, (Jan, 1998) 28–38.
- [16] M. Tomonaga, "Is chimpanzee (*Pan troglodytes*) spatial attention reflexively triggered by gaze cue?," *J Comp Psychol* **121** no. 2, (May, 2007) 156–170.

- [17] R. Adolphs, F. Gosselin, T. W. Buchanan, D. Tranel, P. Schyns, and A. R. Damasio, "A mechanism for impaired fear recognition after amygdala damage," *Nature* **433** no. 7021, (Jan, 2005) 68–72.
- [18] D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cereb. Cortex* **1** no. 1, (1991) 1–47.
- [19] K. Tanaka, H. Saito, Y. Fukada, and M. Moriya, "Coding visual images of objects in the inferotemporal cortex of the macaque monkey," *J. Neurophysiol.* **66** no. 1, (Jul, 1991) 170–189.
- [20] E. Kobatake and K. Tanaka, "Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex," *J. Neurophysiol.* **71** no. 3, (Mar, 1994) 856–867.
- [21] K. Tanaka, "Inferotemporal cortex and object vision," *Annu. Rev. Neurosci.* **19** (1996) 109–139.
- [22] C. S. Sherrington, *The integrative action of the nervous system*. C Scribner and Sons, New York, 1906.
- [23] R. E. Burke and C. Sherrington, "Sir Charles Sherrington's the integrative action of the nervous system: a centenary appreciation," *Brain* **130** no. Pt 4, (Apr, 2007) 887–894.
- [24] J. M. Alonso and Y. Chen, "Receptive field."
http://www.scholarpedia.org/article/Receptive_field.
 Accessed: 2014-12-31.
- [25] H. K. Hartline, "The response of single optic nerve fibers of the vertebrate eye to illumination of the retina," *Am J Physiol* **121** (1938) 400–415.

- [26] J. Allman, F. Miezin, and E. McGuinness, “Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons,” *Annu. Rev. Neurosci.* **8** (1985) 407–430.
- [27] J. R. Cavanaugh, W. Bair, and J. A. Movshon, “Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons,” *J. Neurophysiol.* **88** no. 5, (Nov, 2002) 2547–2556.
- [28] J. R. Cavanaugh, W. Bair, and J. A. Movshon, “Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons,” *J. Neurophysiol.* **88** no. 5, (Nov, 2002) 2530–2546.
- [29] W. E. Vinje and J. L. Gallant, “Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1,” *J. Neurosci.* **22** no. 7, (Apr, 2002) 2904–2915.
- [30] M. J. Chacron, B. Doiron, L. Maler, A. Longtin, and J. Bastian, “Non-classical receptive field mediates switch in a sensory neuron’s frequency tuning,” *Nature* **423** no. 6935, (May, 2003) 77–81.
- [31] C. D. Gilbert, “Laminar differences in receptive field properties of cells in cat primary visual cortex,” *J. Physiol. (Lond.)* **268** no. 2, (Jun, 1977) 391–421.
- [32] M. K. Kapadia, G. Westheimer, and C. D. Gilbert, “Spatial distribution of contextual interactions in primary visual cortex and in visual perception,” *J. Neurophysiol.* **84** no. 4, (Oct, 2000) 2048–2062.
- [33] H. Zhou, H. S. Friedman, and R. von der Heydt, “Coding of border ownership in monkey visual cortex,” *J. Neurosci.* **20** no. 17, (Sep, 2000) 6594–6611.

- [34] W. Li and C. D. Gilbert, “Global contour saliency and local colinear interactions,” *J. Neurophysiol.* **88** no. 5, (Nov, 2002) 2846–2856.
- [35] U. Ernst, K. Pawelzik, F. Wolf, and T. Geisel, “Theory of non-classical receptive field phenomena in the visual cortex,” *Neurocomputing* **2627** no. 0, (1999) 367 – 374. <http://www.sciencedirect.com/science/article/pii/S0925231299000260>.
- [36] T. J. Sullivan and V. R. de Sa, “A model of surround suppression through cortical feedback,” *Neural Netw* **19** no. 5, (Jun, 2006) 564–572.
- [37] C. D. Gilbert and W. Li, “Top-down influences on visual processing,” *Nat. Rev. Neurosci.* **14** no. 5, (May, 2013) 350–363.
- [38] H. Adesnik, W. Bruns, H. Taniguchi, Z. J. Huang, and M. Scanziani, “A neural circuit for spatial summation in visual cortex,” *Nature* **490** no. 7419, (Oct., 2012) 226–231.
<http://www.nature.com/doifinder/10.1038/nature11526>.
- [39] H. Nienborg, A. Hasenstaub, I. Nauhaus, H. Taniguchi, Z. J. Huang, and E. M. Callaway, “Contrast Dependence and Differential Contributions from Somatostatin- and Parvalbumin-Expressing Neurons to Spatial Integration in Mouse V1,” *The Journal of Neuroscience* **33** no. 27, (July, 2013) 11145–11154.
<http://www.jneurosci.org/content/33/27/11145>.
- [40] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” *J. Physiol. (Lond.)* **160** (Jan, 1962) 106–154.
- [41] D. H. HUBEL and T. N. WIESEL, “RECEPTIVE FIELDS AND FUNCTIONAL ARCHITECTURE IN TWO NONSTRIATE VISUAL

- AREAS (18 AND 19) OF THE CAT,” *J. Neurophysiol.* **28** (Mar, 1965) 229–289.
- [42] D. H. Hubel and T. N. Wiesel, “Receptive fields and functional architecture of monkey striate cortex,” *J. Physiol. (Lond.)* **195** no. 1, (Mar, 1968) 215–243.
- [43] M. Riesenhuber and T. Poggio, “Hierarchical models of object recognition in cortex,” *Nat Neurosci* **2** no. 11, (1999) 1019–1025.
- [44] M. Riesenhuber and T. Poggio, “Models of object recognition,” *Nat. Neurosci.* **3 Suppl** (Nov, 2000) 1199–1204.
- [45] T. Serre, A. Oliva, and T. Poggio, “A feedforward architecture accounts for rapid categorization,” *Proc. Natl. Acad. Sci. U.S.A.* **104** no. 15, (Apr, 2007) 6424–6429.
- [46] C. G. Gross, C. E. Rocha-Miranda, and D. B. Bender, “Visual properties of neurons in inferotemporal cortex of the Macaque,” *J. Neurophysiol.* **35** no. 1, (Jan, 1972) 96–111.
- [47] C. G. Gross and J. Sargent, “Face recognition,” *Curr. Opin. Neurobiol.* **2** no. 2, (Apr, 1992) 156–161.
- [48] Y. Miyashita and H. S. Chang, “Neuronal correlate of pictorial short-term memory in the primate temporal cortex,” *Nature* **331** no. 6151, (Jan, 1988) 68–70.
- [49] Y. Sugase, S. Yamane, S. Ueno, and K. Kawano, “Global and fine information coded by single neurons in the temporal visual cortex,” *Nature* **400** no. 6747, (Aug, 1999) 869–873.

- [50] M. P. Young and S. Yamane, “Sparse population coding of faces in the inferotemporal cortex,” *Science* **256** no. 5061, (May, 1992) 1327–1331.
- [51] N. K. Logothetis and D. L. Sheinberg, “Visual object recognition,” *Annu. Rev. Neurosci.* **19** (1996) 577–621.
- [52] N. K. Logothetis, J. Pauls, and T. Poggio, “Shape representation in the inferior temporal cortex of monkeys,” *Curr. Biol.* **5** no. 5, (May, 1995) 552–563.
- [53] T. Sato, “Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques,” *Exp Brain Res* **77** no. 1, (1989) 23–30.
- [54] C. Bruce, R. Desimone, and C. G. Gross, “Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque,” *J. Neurophysiol.* **46** no. 2, (Aug, 1981) 369–384.
- [55] D. I. Perrett, P. A. Smith, D. D. Potter, A. J. Mistlin, A. S. Head, A. D. Milner, and M. A. Jeeves, “Visual cells in the temporal cortex sensitive to face view and gaze direction,” *Proc. R. Soc. Lond., B, Biol. Sci.* **223** no. 1232, (Jan, 1985) 293–317.
- [56] K. Nakamura and K. Kubota, “The primate temporal pole: its putative role in object recognition and memory,” *Behav. Brain Res.* **77** no. 1-2, (May, 1996) 53–77.
- [57] E. T. Rolls, “Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces,” *Hum Neurobiol* **3** no. 4, (1984) 209–222.

- [58] C. M. Leonard, E. T. Rolls, F. A. Wilson, and G. C. Baylis, “Neurons in the amygdala of the monkey with responses selective for faces,” *Behav. Brain Res.* **15** no. 2, (Apr, 1985) 159–176.
- [59] J. E. LeDoux, “Emotional memory systems in the brain,” *Behav. Brain Res.* **58** no. 1-2, (Dec, 1993) 69–79.
- [60] S. Wang, O. Tudusciuc, A. N. Mamelak, I. B. Ross, R. Adolphs, and U. Rutishauser, “Neurons in the human amygdala selective for perceived emotion,” *Proc. Natl. Acad. Sci. U.S.A.* **111** no. 30, (Jul, 2014) E3110–3119.
- [61] I. Fried, K. A. MacDonald, and C. L. Wilson, “Single neuron activity in human hippocampus and amygdala during recognition of faces and objects,” *Neuron* **18** no. 5, (May, 1997) 753–765.
- [62] M. N. Nguyen, E. Hori, J. Matsumoto, A. H. Tran, T. Ono, and H. Nishijo, “Neuronal responses to face-like stimuli in the monkey pulvinar,” *Eur. J. Neurosci.* **37** no. 1, (Jan, 2013) 35–51.
- [63] G. Kreiman, C. Koch, and I. Fried, “Category-specific visual responses of single neurons in the human medial temporal lobe,” *Nat. Neurosci.* **3** no. 9, (Sep, 2000) 946–953.
- [64] R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried, “Invariant visual representation by single neurons in the human brain,” *Nature* **435** no. 7045, (Jun, 2005) 1102–1107.
- [65] E. T. Rolls, H. D. Critchley, A. S. Browning, and K. Inoue, “Face-selective and auditory neurons in the primate orbitofrontal cortex,” *Exp Brain Res* **170** no. 1, (Mar, 2006) 74–87.

- [66] D. Y. Tsao and M. S. Livingstone, “Mechanisms of face perception,” *Annu. Rev. Neurosci.* **31** (2008) 411–437.
- [67] D. Y. Tsao, W. A. Freiwald, R. B. H. Tootell, and M. S. Livingstone, “A cortical region consisting entirely of face-selective cells,” *Science* **311** no. 5761, (2, 2006) 670–4.
- [68] G. C. Baylis, E. T. Rolls, and C. M. Leonard, “Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey,” *Brain Res.* **342** no. 1, (Sep, 1985) 91–102.
- [69] D. I. Perrett, P. A. Smith, D. D. Potter, A. J. Mistlin, A. S. Head, A. D. Milner, and M. A. Jeeves, “Visual cells in the temporal cortex sensitive to face view and gaze direction,” *Proc. R. Soc. Lond., B, Biol. Sci.* **223** no. 1232, (Jan, 1985) 293–317.
- [70] N. Kanwisher, “Functional specificity in the human brain: a window into the functional architecture of the mind,” *Proc. Natl. Acad. Sci. U.S.A.* **107** no. 25, (Jun, 2010) 11163–11170.
- [71] S. M. Zeki, “Functional specialisation in the visual cortex of the rhesus monkey,” *Nature* **274** no. 5670, (Aug, 1978) 423–428.
- [72] J. S. Bowers, “On the biological plausibility of grandmother cells: implications for neural network theories in psychology and neuroscience,” *Psychol Rev* **116** no. 1, (Jan, 2009) 220–251.
- [73] D. Perrett, E. Rolls, and W. Caan, “Visual neurones responsive to faces in the monkey temporal cortex,” *Experimental Brain Research* **47** no. 3, (1982) 329–342. <http://dx.doi.org/10.1007/BF00239352>.

- [74] R. Desimone, T. D. Albright, C. G. Gross, and C. Bruce, "Stimulus-selective properties of inferior temporal neurons in the macaque," *J. Neurosci.* **4** no. 8, (Aug, 1984) 2051–2062.
- [75] S. Yamane, S. Kaji, and K. Kawano, "What facial features activate face neurons in the inferotemporal cortex of the monkey?," *Exp Brain Res* **73** no. 1, (1988) 209–214.
- [76] M. E. Hasselmo, E. T. Rolls, and G. C. Baylis, "The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey," *Behav. Brain Res.* **32** no. 3, (Apr, 1989) 203–218.
- [77] M. H. Harries and D. I. Perrett, "Visual processing of faces in temporal cortex: physiological evidence for a modular organization and possible anatomical correlates," *J Cogn Neurosci* **3** no. 1, (1991) 9–24.
- [78] P. Foldiak, D. Xiao, C. Keysers, R. Edwards, and D. I. Perrett, "Rapid serial visual presentation for the determination of neural selectivity in area STSa," *Prog. Brain Res.* **144** (2004) 107–116.
- [79] S. Eifuku, W. C. De Souza, R. Tamura, H. Nishijo, and T. Ono, "Neuronal correlates of face identification in the monkey anterior temporal cortical areas," *J. Neurophysiol.* **91** no. 1, (Jan, 2004) 358–371.
- [80] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," *Proc. Natl. Acad. Sci. U.S.A.* **87** no. 24, (Dec, 1990) 9868–9872.
- [81] S. A. Huettel, A. W. Song, and G. McCarthy, *Functional Magnetic Resonance Imaging (2 ed.)*, Massachusetts. Sinauer, 2009.

- [82] J. Sergent, S. Ohta, and B. MacDonald, “Functional neuroanatomy of face and object processing. A positron emission tomography study,” *Brain* **115 Pt 1** (Feb, 1992) 15–36.
- [83] N. Kanwisher, J. McDermott, and M. M. Chun, “The fusiform face area: a module in human extrastriate cortex specialized for face perception,” *J. Neurosci.* **17** no. 11, (Jun, 1997) 4302–4311.
- [84] N. Kanwisher and G. Yovel, “The fusiform face area: a cortical region specialized for the perception of faces,” *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **361** no. 1476, (Dec, 2006) 2109–2128.
- [85] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini, “Distributed and overlapping representations of faces and objects in ventral temporal cortex,” *Science* **293** no. 5539, (Sep, 2001) 2425–2430.
- [86] P. E. Downing, Y. Jiang, M. Shuman, and N. Kanwisher, “A cortical area selective for visual processing of the human body,” *Science* **293** no. 5539, (Sep, 2001) 2470–2473.
- [87] D. Y. Tsao, S. Moeller, and W. A. Freiwald, “Comparing face patch systems in macaques and humans,” *Proceedings of the National Academy of Sciences* **105** no. 49, (Dec., 2008) 19514–19519.
<http://www.pnas.org/content/105/49/19514>.
- [88] S. Moeller, W. A. Freiwald, and D. Y. Tsao, “Patches with links: a unified system for processing faces in the macaque temporal lobe,” *Science* **320** no. 5881, (Jun, 2008) 1355–1359.

- [89] W. A. Freiwald and D. Y. Tsao, “Functional compartmentalization and viewpoint generalization within the macaque face-processing system.,” *Science* **330** no. 6005, (11, 2010) 845–51.
- [90] E. B. Issa and J. J. DiCarlo, “Precedence of the eye region in neural processing of faces,” *J. Neurosci.* **32** no. 47, (Nov, 2012) 16666–16682.
- [91] G. Wang, K. Tanaka, and M. Tanifuji, “Optical imaging of functional organization in the monkey inferotemporal cortex,” *Science* **272** no. 5268, (Jun, 1996) 1665–1668.
- [92] E. T. Rolls and M. J. Tovee, “Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex.,” *J Neurophysiol* **73** no. 2, (1995) 713–726.
- [93] N. Liu, N. Kriegeskorte, M. Mur, F. Hadj-Bouziane, W.-M. M. Luh, R. B. H. Tootell, and L. G. Ungerleider, “Intrinsic structure of visual exemplar and category representations in macaque brain.,” *J Neurosci* **33** no. 28, (7, 2013) 11346–11360.
- [94] P. Sinha, “Recognizing complex patterns,” *Nat. Neurosci.* **5 Suppl** (Nov, 2002) 1093–1097.
- [95] K. H. Britten and H. W. Heuer, “Spatial summation in the receptive fields of MT neurons,” *The Journal of Neuroscience* **19** no. 12, (June, 1999) 5074–5084. <http://www.jneurosci.org/content/19/12/5074>.
- [96] J. H. Reynolds, L. Chelazzi, and R. Desimone, “Competitive mechanisms subserve attention in macaque areas V2 and V4,” *J. Neurosci.* **19** no. 5, (Mar, 1999) 1736–1753. [PubMed:10024360].
- [97] J. Moran and R. Desimone, “Selective attention gates visual processing in the extrastriate cortex,” *Science* **229** no. 4715, (Aug, 1985) 782–784.

- [98] S. Treue and J. H. Maunsell, “Attentional modulation of visual motion processing in cortical areas MT and MST,” *Nature* **382** no. 6591, (Aug, 1996) 539–541.
- [99] R. Desimone and J. Duncan, “Neural mechanisms of selective visual attention,” *Annu. Rev. Neurosci.* **18** (1995) 193–222.
- [100] D. J. Heeger, “Normalization of cell responses in cat striate cortex,” *Visual Neuroscience* (1992) .
- [101] M. Carandini, D. J. Heeger, and J. A. Movshon, “Linearity and normalization in simple cells of the macaque primary visual cortex,” *J Neurosci* (1997) .
- [102] E. P. Simoncelli and D. J. Heeger, “A model of neuronal responses in visual area MT,” *Vision Research* **38** no. 5, (Mar., 1998) 743–761.
- [103] M. Carandini and D. J. Heeger, “Normalization as a canonical neural computation.,” *Nat Rev Neurosci* **13** no. 1, (1, 2012) 51–62.
- [104] S. P. Mysore and E. I. Knudsen, “Reciprocal inhibition of inhibition: a circuit motif for flexible categorization in stimulus selection,” *Neuron* **73** no. 1, (Jan, 2012) 193–205.
- [105] S. X. Luo, R. Axel, and L. F. Abbott, “Generating sparse and selective third-order responses in the olfactory system of the fly,” *Proc Natl Acad Sci U S A* **107** no. 23, (6, 2010) 10713–8.
- [106] S. R. Olsen, V. Bhandawat, and R. I. Wilson, “Divisive normalization in olfactory population codes,” *Neuron* **66** no. 2, (4, 2010) 287–99.

- [107] D. Zoccolan, D. D. Cox, and J. J. DiCarlo, “Multiple object response normalization in monkey inferotemporal cortex.,” *J Neurosci* **25** no. 36, (9, 2005) 8150–64.
- [108] A. M. Ni, S. Ray, and J. H. R. Maunsell, “Tuned normalization explains the size of attention modulations.,” *Neuron* **73** no. 4, (2, 2012) 803–13.
- [109] J. Lee and J. H. R. Maunsell, “A normalization model of attentional modulation of single unit responses.,” *PLoS One* **4** no. 2, (2009) e4651.
- [110] T. Ohshiro, D. E. Angelaki, and G. C. DeAngelis, “A normalization model of multisensory integration,” *Nature Neuroscience* **14** no. 6, (June, 2011) 775–782.
- [111] J. H. Reynolds and D. J. Heeger, “The normalization model of attention,” *Neuron* **61** no. 2, (Jan, 2009) 168–185.
- [112] L. Busse, A. R. Wade, and M. Carandini, “Representation of concurrent stimuli by population activity in visual cortex,” *Neuron* **64** no. 6, (Dec, 2009) 931–942.
- [113] S. P. MacEvoy, T. R. Tucker, and D. Fitzpatrick, “A precise form of divisive suppression supports population coding in the primary visual cortex,” *Nature Neuroscience* **12** no. 5, (May, 2009) 637–645. <http://www.nature.com/neuro/journal/v12/n5/full/nn.2310.html>.
- [114] T. K. Sato, M. Häusser, and M. Carandini, “Distal connectivity causes summation and division across mouse visual cortex.,” *Nat Neurosci* **17** no. 1, (1, 2014) 30–2.
- [115] J. Reynolds, “Modeling optogenetic manipulation of neural circuits in macaque visual cortex.” Cold Spring Harbor Symposium on Quantitative Biology: Cognition, Cold Spring Harbor, NY., 2014.

- [116] G. G. Gregoriou, S. J. Gotts, and R. Desimone, “Cell-type-specific synchronization of neural activity in FEF with V4 during attention,” *Neuron* **73** no. 3, (Feb, 2012) 581–594.
- [117] D. Zoccolan, M. Kouh, T. Poggio, and J. J. DiCarlo, “Trade-off between object selectivity and tolerance in monkey inferotemporal cortex,” *J. Neurosci.* **27** no. 45, (Nov, 2007) 12292–12307.
- [118] E. T. Rolls, N. C. Aggelopoulos, and F. Zheng, “The receptive fields of inferior temporal cortex neurons in natural scenes,” *J. Neurosci.* **23** no. 1, (Jan, 2003) 339–348.
- [119] H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster, “Inhibitory stabilization of the cortical network underlies visual surround suppression,” *Neuron* **62** no. 4, (May, 2009) 578–592.
- [120] D. B. Rubin, S. D. Van Hooser, and K. D. Miller, “The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex,” *Neuron* **85** no. 2, (Jan., 2015) 402–417.
- [121] M. Turatto, M. Valsecchi, A. E. Seiffert, and A. Caramazza, “On the speed of pop-out in feature search,” *J Exp Psychol Hum Percept Perform* **36** no. 5, (Oct, 2010) 1145–1152.
- [122] O. Hershler and S. Hochstein, “At first sight: a high-level pop out effect for faces,” *Vision Res.* **45** no. 13, (Jun, 2005) 1707–1724.
- [123] A. M. Treisman and G. Gelade, “A feature-integration theory of attention,” *Cogn Psychol* **12** no. 1, (Jan, 1980) 97–136.
- [124] A. Treisman, “Perceptual grouping and attention in visual search for features and for objects,” *J Exp Psychol Hum Percept Perform* **8** no. 2, (Apr, 1982) 194–214.

- [125] H. C. Nothdurft, “Faces and facial expressions do not pop out,” *Perception* **22** no. 11, (1993) 1287–1298.
- [126] A. Treisman and J. Souther, “Search asymmetry: A diagnostic for preattentive processing of separable features,” *Journal of Experimental Psychology: General* **114** no. 3, (1985) 285–310.
- [127] H. C. Nothdurft, J. L. Gallant, and D. C. Van Essen, “Response modulation by texture surround in primate area V1: correlates of “popout” under anesthesia,” *Vis. Neurosci.* **16** no. 1, (1999) 15–34.
- [128] S. Kastner, H. C. Nothdurft, and I. N. Pigarev, “Neuronal responses to orientation and motion contrast in cat striate cortex,” *Vis. Neurosci.* **16** no. 3, (1999) 587–600.
- [129] D. M. Beck and S. Kastner, “Stimulus context modulates competition in human extrastriate cortex,” *Nat. Neurosci.* **8** no. 8, (Aug, 2005) 1110–1116.
- [130] F. Gentile and B. Jansma, “Neural competition through visual similarity in face selection,” *Brain Research* **1351** (Sept., 2010) 172–184. <http://linkinghub.elsevier.com/retrieve/pii/S0006899310014423>.
- [131] K. Nagy, M. W. Greenlee, and G. Kovcs, “Sensory competition in the face processing areas of the human brain,” *PloS one* **6** no. 9, (2011) e24450. <http://dx.plos.org/10.1371/journal.pone.0024450>.
- [132] D. M. Levi, S. Hariharan, and S. A. Klein, “Suppressive and facilitatory spatial interactions in peripheral vision: peripheral crowding is neither size invariant nor simple contrast masking,” *J Vis* **2** no. 2, (2002) 167–177.

- [133] D. M. Levi, S. Hariharan, and S. A. Klein, “Suppressive and facilitatory spatial interactions in amblyopic vision,” *Vision Res.* **42** no. 11, (May, 2002) 1379–1394.
- [134] D. M. Levi, “Crowding—an essential bottleneck for object recognition: a mini-review,” *Vision Res.* **48** no. 5, (Feb, 2008) 635–654.
- [135] D. Whitney and D. M. Levi, “Visual crowding: a fundamental limit on conscious perception and object recognition.,” *Trends Cogn Sci* **15** no. 4, (2011) 160–168.
- [136] H. Bouma, “Interaction effects in parafoveal letter recognition,” *Nature* (1970) .
- [137] A. Toet and D. M. Levi, “The two-dimensional shape of spatial interaction zones in the parafovea,” *Vision Res.* **32** no. 7, (Jul, 1992) 1349–1357.
- [138] C. Feng, Y. Jiang, and S. He, “Horizontal and vertical asymmetry in visual spatial crowding effects,” *J Vis* **7** no. 2, (2007) 1–10.
- [139] L. Parkes, J. Lund, A. Angelucci, J. A. Solomon, and M. Morgan, “Compulsory averaging of crowded orientation signals in human vision,” *Nat. Neurosci.* **4** no. 7, (Jul, 2001) 739–744.
- [140] D. G. Pelli, M. Palomares, and N. J. Majaj, “Crowding is unlike ordinary masking: distinguishing feature integration from detection,” *J Vis* **4** no. 12, (Dec, 2004) 1136–1169.
- [141] J. A. Greenwood, P. J. Bex, and S. C. Dakin, “Positional averaging explains crowding with letter-like stimuli,” *Proc. Natl. Acad. Sci. U.S.A.* **106** no. 31, (Aug, 2009) 13130–13135.

- [142] B. Balas, L. Nakano, and R. Rosenholtz, “A summary-statistic representation in peripheral vision explains visual crowding,” *J Vis* **9** no. 12, (2009) 1–18.
- [143] J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *Int’l Journal of Computer Vision* **40** no. 1, (December, 2000) 49–71.
- [144] G. Wallis and E. T. Rolls, “Invariant face and object recognition in the visual system,” *Prog. Neurobiol.* **51** no. 2, (Feb, 1997) 167–194.
- [145] F. Wilkinson, H. R. Wilson, and D. Ellemberg, “Lateral interactions in peripherally viewed texture arrays,” *J Opt Soc Am A Opt Image Sci Vis* **14** no. 9, (Sep, 1997) 2057–2068.
- [146] R. van den Berg, J. B. Roerdink, and F. W. Cornelissen, “A neurophysiologically plausible population code model for feature integration explains visual crowding,” *PLoS Comput. Biol.* **6** no. 1, (Jan, 2010) e1000646.
- [147] E. R. Kandel, J. H. Schwartz, and T. M. Jessel, *Essentials of neural science and behavior*. McGraw-Hill, Appleton & Lange, 1996.
- [148] J. Freeman and E. P. Simoncelli, “Metamers of the ventral stream,” *Nat. Neurosci.* **14** no. 9, (Sep, 2011) 1195–1201.
- [149] R. Rosenholtz, J. Huang, A. Raj, B. J. Balas, and L. Ilie, “A summary statistic representation in peripheral vision explains visual search,” *J Vis* **12** no. 4, (2012) .
- [150] G. Wolford, “Perturbation model for letter identification,” *Psychol Rev* **82** no. 3, (May, 1975) 184–199.

- [151] H. Strasburger and M. Malania, “Source confusion is a major cause of crowding,” *J Vis* **13** no. 1, (2013) .
- [152] E. F. Ester, D. Klee, and E. Awh, “Visual crowding cannot be wholly explained by feature pooling,” *J Exp Psychol Hum Percept Perform* **40** no. 3, (Jun, 2014) 1022–1033.
- [153] P. Dayan and J. A. Solomon, “Selective Bayes: attentional load and crowding,” *Vision Res.* **50** no. 22, (Oct, 2010) 2248–2260.
- [154] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, 2005.
- [155] D. Y. Tsao, N. Schweers, S. Moeller, and W. A. Freiwald, “Patches of face-selective cortex in the macaque frontal lobe,” *Nat. Neurosci.* **11** no. 8, (Aug, 2008) 877–879.
- [156] D. C. Van Essen, H. A. Drury, J. Dickson, J. Harwell, D. Hanlon, and C. H. Anderson, “An integrated software suite for surface-based analyses of cerebral cortex,” *J Am Med Inform Assoc* (2001) .
- [157] A. Rosset, L. Spadola, and O. Ratib, “OsiriX: an open-source software for navigating in multidimensional DICOM images,” *J Digit Imaging* **17** no. 3, (Sep, 2004) 205–216.
- [158] S. Ohayon and D. Y. Tsao, “Mr-guided stereotactic navigation,” *J Neurosci Methods* **204** no. 2, (3, 2012) 389–97.
- [159] C. F. Crist, D. S. Yamasaki, H. Komatsu, and R. H. Wurtz, “A grid system and a microsyringe for single cell recording,” *J Neurosci Methods* (1988) .

- [160] M. G. Stokes, M. Kusunoki, N. Sigala, H. Nili, D. Gaffan, and J. Duncan, “Dynamic coding for cognitive control in prefrontal cortex,” *Neuron* **78** no. 2, (Apr, 2013) 364–375.
- [161] K. Hasegawa, *Honto ni wakaru taehenryou kaiseki. [Introduction to multivariate data analysis.]*. Kyoritsu Syuppan, 1998.
- [162] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley-Interscience, 2001.
- [163] A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” *Neural Comput* **7** no. 6, (Nov, 1995) 1129–1159.
- [164] T. Kolenda, S. Sigurdsson, O. Winther, L. K. Hansen, and J. Larsen, “DTU:toolbox. ISP Group, Informatics and Mathematical Modeling, Technical University of Denmark.,” *[Computer program]* (2002) .
- [165] C. K. Machens, “Demixing population activity in higher cortical areas,” *Front Comput Neurosci* **4** (2010) 126.
- [166] E. M. Meyers, “The neural decoding toolbox.,” *Front Neuroinform* **7** (2013) 8.
- [167] P. Grimaldi, K. S. Saleem, and D. Y. Tsao, “Aanatomical connections of functionally defined anterior face patches in the macaque monkey.” Society of Neuroscience annual meeting. Poster session on Neural processing of faces and bodies. New Orleans, LA., 2012.
- [168] E. M. Meyers, M. Borzello, W. A. Freiwald, and D. Y. Tsao, “Decoding what types of information are in the macaque face patch system.” Society of Neuroscience annual meeting. Nanosymposium on Representation of faces and bodies. Washington, DC., 2014.

- [169] K. Mirpour and H. Esteky, “State-dependent effects of stimulus presentation duration on the temporal dynamics of neural responses in the inferotemporal cortex of macaque monkeys,” *J. Neurophysiol.* **102** no. 3, (Sep, 2009) 1790–1800.
- [170] R. N. Sachdev and K. C. Catania, “Effects of stimulus duration on neuronal response properties in the somatosensory cortex of the star-nosed mole,” *Somatosens Mot Res* **19** no. 4, (2002) 272–278.
- [171] B. C. Motter, “Modulation of transient and sustained response components of V4 neurons by temporal crowding in flashed stimulus sequences,” *J. Neurosci.* **26** no. 38, (Sep, 2006) 9683–9694.
- [172] M. Rigotti, O. Barak, M. R. Warden, X. J. Wang, N. D. Daw, E. K. Miller, and S. Fusi, “The importance of mixed selectivity in complex cognitive tasks,” *Nature* **497** no. 7451, (May, 2013) 585–590.
- [173] M. Pagan, L. S. Urban, M. P. Wohl, and N. C. Rust, “Signals in inferotemporal and perirhinal cortex suggest an untangling of visual target information,” *Nat. Neurosci.* **16** no. 8, (Aug, 2013) 1132–1139.
- [174] G. Mitchison, “Neuronal branching patterns and the economy of cortical wiring,” *Proc. Biol. Sci.* **245** no. 1313, (Aug, 1991) 151–158.
- [175] O. Sporns, “The non-random brain: efficiency, economy, and complex dynamics,” *Front Comput Neurosci* **5** (2011) 5.
- [176] J. M. Wallace and B. S. Tjan, “Object crowding.,” *J Vis* **11** no. 6, (2011) .
- [177] M. R. Cohen and J. H. Maunsell, “Attention improves performance primarily by reducing interneuronal correlations,” *Nat. Neurosci.* **12** no. 12, (Dec, 2009) 1594–1600.

- [178] G. Mongillo, O. Barak, and M. Tsodyks, “Synaptic theory of working memory,” *Science* **319** no. 5869, (Mar, 2008) 1543–1546.
- [179] S. L. Brincat and C. E. Connor, “Dynamic shape synthesis in posterior inferotemporal cortex,” *Neuron* **49** no. 1, (Jan, 2006) 17–24.
- [180] M. Pagan and N. C. Rust, “Dynamic Target Match Signals in Perirhinal Cortex Can Be Explained by Instantaneous Computations That Act on Dynamic Input from Inferotemporal Cortex,” *The Journal of Neuroscience* **34** no. 33, (Aug, 2014) 11067–11084.
- [181] J. Schmahmann and D. Pandya, *Fiber Pathways of the Brain*. Oxford University Press, 2009.
- [182] M. H. Johnson, “Subcortical face processing,” *Nat. Rev. Neurosci.* **6** no. 10, (Oct, 2005) 766–774.
- [183] J. I. Gold and M. N. Shadlen, “The neural basis of decision making,” *Annu. Rev. Neurosci.* **30** (2007) 535–574.
- [184] J. R. Busemeyer and J. T. Townsend, “Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment,” *Psychol Rev* **100** no. 3, (Jul, 1993) 432–459.
- [185] P. L. Smith, “Stochastic Dynamic Models of Response Time and Accuracy: A Foundational Primer,” *J Math Psychol* **44** no. 3, (Sep, 2000) 408–463.
- [186] W. Vanduffel, D. Fize, J. B. Mandeville, K. Nelissen, P. Van Hecke, B. R. Rosen, R. B. H. Tootell, and G. A. Orban, “Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys,” *Neuron* **32** no. 4, (Nov., 2001) 565–577.
<http://www.cell.com/article/S0896627301005025/abstract>.

- [187] P. Polosecki, S. Moeller, N. Schweers, L. M. Romanski, D. Y. Tsao, and W. A. Freiwald, “Faces in motion: Selectivity of macaque and human face processing areas for dynamic stimuli,” *The Journal of Neuroscience* **33** no. 29, (July, 2013) 11768–11773.
<http://www.jneurosci.org/content/33/29/11768>.
- [188] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, “Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering,” *Neural Comput* **16** no. 8, (Aug, 2004) 1661–1687.
- [189] P. M. Daniel and D. Whitteridge, “The representation of the visual field on the cerebral cortex in monkeys,” *The Journal of physiology* **159** no. 2, (1961) 203–221. <http://onlinelibrary.wiley.com/doi/10.1113/jphysiol.1961.sp006803/abstract>.
- [190] “Neuroanatomy and neuroembryology.” <http://www.rci.rutgers.edu/~uzwiak/AnatPhys/APFallLect19.html>.
Accessed: 2014-12-25.
- [191] *MATLAB 8.4.0*. 2014. The MathWorks, Inc., Natick, Massachusetts, United States.
- [192] “Mathematics stack exchange.”
<http://math.stackexchange.com/questions/3869/what-is-the-intuitive-relationship-between-svd-and-pca>.
Accessed: 2015-1-6.
- [193] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, “Singular Value Decomposition and Principal Component Analysis,” *ArXiv Physics e-prints* (Aug., 2002) , [physics/0208101](https://arxiv.org/abs/physics/0208101).

- [194] J. Shlens, “A tutorial on principal component analysis,” *CoRR* **abs/1404.1100** (2014) . <http://arxiv.org/abs/1404.1100>.