Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :
Biao Han

Le jeudi 7 avril 2016

Titre :

Predictive coding: its spike-time based neuronal implementation and its relationship with perception and oscillations

---

ED CLESCO : Neurosciences

Unité de recherche :

Centre de Recherche Cerveau & Cognition - CNRS UMR5549

Directeur(s) de Thèse :

Rufin VanRullen

Rapporteurs :

Christopher Summerfield
Srdjan Ostojic

Autre(s) membre(s) du jury :

Simon Thorpe

“If the brain were simple enough for us to understand, we would be too simple to understand it.”

Ken Hill

“What we know is enough to convince anyone that the brain, though complicated, works in a way that will probably someday be understood —and that the answers will not be so complicated that they can be understood only by people with degrees in computer science or particle physics.”

David Hubel

Acknowledgements

From Bachelor in Economics, to Master in Communication Engineering, and then PhD in Neuroscience, I have quite an interesting and unique journey in my academic life. I chose the research of neuroscience because it asks one ultimate question: how our brain works? With modern technology in neuroscience and the achievements in computer science, I believe that this problem is solvable in our lifetime. I feel really lucky that I could join CerCo. I often got excited and proud when I read something that I consider important and the work was done by the people from our lab. In my three years' study in CerCo, I would like to thank the following people:

Rufin, you are one of the smartest person I have ever met. I think I will miss the discussion and debate that I had with you for the rest of my life. Thank you so much for helping me with the endless submission and resubmission. I enjoyed the life in our team.

Doug, I learned almost all my neurophysiology knowledge from you. Thank you so much for sparing no pains to help me with all of the neuroscience questions. Your help is not limited to science, but also in my everyday life. I am grateful for all of your help and I cherish my friendship with you.

Simon, Lionel, Jean-Michel, Pascal Girard and Pascal Barone, I am proud of being in a lab with great researchers like you. Thank you very much for helping me with my queries about the knowledge of neuroscience. Zoé and Claire, thank you for helping me with the miscellaneous administration affairs.

Benedikt, Marie, Sasskia, Rodika, Marina, Mehdi, Anne-Claire, Seb, Rasa Diego, Marcello and others, I enjoyed my life with you guys and I really wish see you guys in my future personal and academic life. For all the French speakers, thank you a lot for your help in French, that's the only reason that I can survive here.

Thanks mom and dad for the long-term support. Lu, I am sorry that I could not be around you. Thank you for being a great companion for all the days and nights to me and we will be together soon. I love you.

Abstract

In this thesis, we investigated predictive coding and its relationship with perception and oscillations. We first reviewed my current understanding about facts of neuron and neocortex and state-of-the-arts of predictive coding in the introduction. In the main chapters, firstly, we proposed the idea that correlated spike times create selective inhibition in a non-selective excitatory feedback network in a theoretical study. Then, we showed the perceptual effect of predictive coding: shape perception enhances perceived contrast. At last, we showed that predictive coding can use oscillations with different frequencies for feedforward and feedback. This thesis provided an innovative and viable neuronal mechanism for predictive coding and empirical evidence for excitatory predictive feedback and the close relationship between the predictive coding and oscillations.

Keywords: predictive coding, temporal coding, excitatory feedback, oscillations

Résumé

Dans cette thèse, nous avons étudié le codage prédictif and sa relation avec la perception et les oscillations. Nous avons, dans l'introduction, fait une revue des connaissances sur les neurones et le néocortex et un état de l'art du codage prédictif. Dans les chapitres principaux, nous avons tout d'abord, proposé l'idée, au travers d'une étude théorique, que la temporalité de la décharge crée une inhibition sélective dans les réseaux excitateurs non-sélectifs rétroactifs. Ensuite, nous avons montré les effets perceptuels du codage prédictif: la perception de la forme améliore la perception du contraste. Enfin, nous avons montré que le codage prédictif peut utiliser des oscillations dans différentes bandes de fréquences pour transmettre les informations en avant et en rétroaction. Cette thèse a fourni un mécanisme neuronal viable et innovant pour le codage prédictif soutenu par des données empiriques démontrant des prédictions rétroactives excitatrices et une relation forte entre codage prédictif et oscillations.

Mots-clés: codage prédictif, codage temporel, rétroactivité excitatoire, oscillation.

Résumé substantiel

Dans cette thèse, nous avons étudié le codage prédictif et sa relation avec la perception et les oscillations.

En introduction, il a été fait une revue des données empiriques qui nous semblaient fondamentales et universelles et un état de l'art des connaissances actuelles sur le codage prédictif.

Dans le chapitre I, nous avons fait un examen théorique du codage prédictif, qui est au cœur de cette thèse. Puisque le modèle du codage prédictif classique n'est pas un modèle neuronal, nous avons proposé un modèle du codage prédictif basé sur la corrélation entre les moments de décharges neuronales. Cette étude a été motivée par les contradictions étonnantes dues à l'inhibition rétroactive: la rétroactivité peut avoir un effet à la fois sélectif et inhibiteur, alors que les connections rétroactives sont divergentes et excitatrices. Dans cette étude, nous avons démontré qu'il est possible de régénérer un effet d'inhibition sélective en utilisant la causalité entre les moments de décharges neuronales des aires supérieures et des aires les plus basses, et de la courbe de réponse entre phase et temporalité de décharge, une propriété de réponse fondamentale des neurones.

Nous avons tout d'abord démontré, dans les simulations, que les neurones des aires les plus basses répondent moins à l'excitation par rétroaction (inhibition relative) quand le moment de décharge est corrélé avec les moments de décharge des neurones actifs dans les aires supérieures. Les mécanismes sous-tendant cet effet sont basés sur les différents déplacements vers l'avant des moments de décharge

neuronale pour différents temps de rétroaction par rapport au dernier moment de décharge du neurone de bas niveau. Les neurones prédictibles (ceux de bas niveau qui entrainent les neurones des aires supérieures) reçoivent des retours d'information juste après leurs dernières décharges, ces derniers ayant, par conséquent un effet très limité sur l'activité des neurones de bas niveau. D'un autre côté, les neurones imprédictibles (les neurones de bas niveaux qui n'entrainent pas les neurones des aires supérieures) reçoivent un déplacement vers l'avant moyen de la temporalité de leurs décharges. Nous avons ensuite montré les quatre facteurs qui peuvent influencer le retour d'information et la sélectivité basé sur le moment de décharge: la force du retour d'information, le délai de conduction axonal, bruit dans le système et la prévisibilité des neurones prédictibles. Nous avons montré que la force de retour de l'information permet de moduler la sélectivité de deux manières: d'une part au travers d'une relation monotone entre la sélectivité et le délai de conduction axonale (délai plus court et effet plus fort), et entre la sélectivité et la prévisibilité (les neurones de bas niveau plus prédictibles créent une sélectivité plus forte). Nous avons aussi montré la forte résistance de ce modèle face au bruit du système. Ensuite, nous avons démontré que la normalisation dans les aires plus basses peut transformer l'inhibition relative en inhibition absolue. Le principe computationnel proposé fourni un mécanisme neuronal viable pour un codage efficace avec une sélectivité basée sur des moments de décharge beaucoup plus flexible que la sélectivité traditionnelle basée sur les poids de connectivité.

Nous avons subséquemment abordé la question du rôle de la plasticité dépendante des moments de décharge des neurones (« Spike-time depedent plasticity », STDP) dans de tels modèles. Nous avons montré que la corrélation entre les moments de décharge neuronale générée

dans le modèle peut bénéficier de la STDP pour augmenter les effets inhibiteurs et sélectifs existants.

Dans le chapitre II, inspiré par les connections excitatrices rétroactives dans le modèle, nous avons employé une approche psychophysique pour évaluer l'effet perceptuel du codage prédictif puisque la majorité des études utilisant l'IRMf ont montré un effet inhibiteur du retour de l'information prédictif.

Pour produire une rétroactivité prédictive, nous avons utilisé des stimuli similaires à ceux utilisés par Murray et collègues : c'est-à-dire des contours de formes en 3D et des versions de lignes aléatoires (Murray et al., 2002). Ces premières peuvent être facilement reconnaissables, et devraient normalement produire plus de rétroaction prédictive que ces dernières. Les deux types de stimuli (3D et lignes aléatoires) étaient montrés simultanément sur des disques gris à droite et à gauche d'un point de fixation sur fond noir. Les sujets avaient pour tâche de comparer la luminance de deux disques (et rapporter quel côté était le plus lumineux). Nous avons obtenu des réponses comportementales de 14 sujets (incluant 2 sujets avec un oculomètre) et nous avons trouvé une réponse comportementale constante montrant que le disque derrière le stimulus en 3D était perçut comme plus lumineux contre un fond noir que le disque gris avec le stimulus composé de lignes aléatoires (sans sens). Puisque des études antérieures ont suggéré une relation monotone entre perception du contraste et activité dans les aires visuelles primaires (Dean, 1981; Boynton et al., 1999), nous interprétons ces résultats comme une preuve que la rétroactivité prédictive a un effet excitateur sur les activités sensorielles comme suggéré par notre modèle.

Nous avons effectué des expériences contrôles pour éliminer trois explications alternatives à nos résultats: un biais attentionnel, des facteurs locaux et un biais de réponse. Les manipulations effectuées pour réaliser les expériences contrôles incluaient le remplacement du point de fixation central par une tâche à forte demande attentionnelle (lettre RSVP), renversement de la polarité du contour des stimuli (de noir à blanc), la modification des instructions de réponse (en demandant "quel disque était plus foncé" au lieu de "quel disque était plus clair?), et en changeant la tâche des sujets (en tâche de perception même/différente luminance en demande "Est-ce que les deux disque avaient la même luminance?"). Ces expériences contrôles ont montré que les explications alternatives de nos résultats peuvent être écartées.

Dans le chapitre III, nous avons décrit une étude sur la relation entre le codage prédictif et les oscillations. Puisque la théorie du codage prédictif suggérait que les interactions entre aires plus basses et aires supérieures étaient de nature itérative, il est intuitif de supposer que le codage prédictif bénéficie des oscillations neuronales et que les prédictions et les erreurs de prédictions pourraient moduler le traitement sensorielle périodiquement. Puisque la phase pourrait refléter l'état de l'oscillation, nous avons étudié la relation entre la phase pré-stimulus (puisqu'il n'y a pas de réinitialisation de celle-ci par le stimulus) et l'effet perceptuel du codage prédictif que nous avons observé dans l'étude précédente.

Nous avons utilisé un paradigme similaire à l'étude précédente en induisant différentes quantités de rétroactivité prédictive (forme 3D ou lignes aléatoires), et nous avons mesuré les effets correspondant sur le jugement de la luminance comme marqueur pour chaque essai de

l'efficacité du codage prédictif en même temps que l'activité EEG était enregistrée. En analysant la relation entre décision après le stimulus et la phase de l'EEG avant le stimulus, qui est un marqueur de la phase présente lors que la prédiction arrive (après apparition de la forme 3D sa représentation dans les aires supérieures est renvoyée vers l'arrière), nous avons trouvé que deux oscillations spontanées avant le stimulus dans différentes régions et fréquences pouvaient fortement influencer le jugement de luminance: les oscillations thêta controlatérales frontales (aires supérieures) et les oscillations béta controlatérales occipitales (aires inférieures). La phase de l'oscillation thêta avant le déclanchement du stimulus pouvait expliquer 14% de la différence de jugement de luminance alors que la phase de l'oscillation beta pouvait en expliquer 19%. Des analyses contrôles ont éliminé la possibilité de contamination de la relation phase-comportement par des activités post-stimuli ou des artefacts oculaires. Ces résultats suggèrent non seulement que le codage prédictif est un processus périodique mais révèlent également deux périodicités avec des sources différentes: le cerveau renvoie les prédictions à une fréquence thêta, et les erreurs de prédiction à une fréquence béta.

Pour conclure, nous avons effectué une discussion générale de cette thèse exposant ses forces et ses faiblesses et les possibilités de développement des thèmes abordés.

Table of Contents

Introduction

The brain is the hardware of our conscious self. We know that we use the brain to do different kinds of things everyday such as reading books, recognizing objects, identifying faces, driving cars, even when we are lying on the beach, we still need the brain to feel the heat from the sunshine on our back. It is amazing that nature itself could build this kind of organ. Even though neuroscience is a young field of research and there are so many questions we cannot answer, we are not absolutely ignorant about the brain.

If we consider the brain as a machine, the fundamental pieces of this machine would be the neurons. The neurons are also called nerve cells, which are just nothing but one special kind of cells. However, one different point between neurons and other cells made neurons special: neurons send information through electrical and chemical signals via synapses. To start to understand the brain, the first thing we should understand is the neuron itself. The physical properties of individual neurons such as size, shape, axon/dendrites number and length are important since they determine physical possibilities of each neuron. Furthermore, different neurons' morphology could be usually related to the types of neurotransmitters released from the neuron's axon, which is important since it is directly linked to the functional role of each neuron.

More importantly, neurons can connect together to form structures. Each neuron has its own dendrites and axons. The principle is simple: pre-synaptic axon contacts the post-synaptic dendrites (sometimes, soma)

to form connections. However, the interconnections between different neurons and different groups of neurons are complicated. We could see this in two ways: if we want to understand the interconnections of one particular neuron, it is dependent on the properties of the surrounding environment of that neuron such as the density, location and surrounding neuron types. If we want to understand the interconnections between two connected neurons, the interconnections are dependent on the relative properties such as the relative hierarchical position (feedforward or feedback connections), relative distance and so on. One can imagine that the special physical properties are not the cause of functional purpose but the results of that. Since we have already observed common features of neurons and their structures, we should be able to infer the functional purpose.

One structure in our brain probably has the most amazing functional purpose: the visual system. We see the world using the visual system, and gain most of the information from the outside world into our mind through this system. In the brain, we know that this system starts from retina (remember that the retina is a part of the brain), and there is a classical pathway to project the information to the back of the brain to the primary visual cortex, and then project back to the frontal area for decision and control (dorsal stream) or to the hippocampus for memory (ventral stream). Since the visual system is also formed by neurons, it should inherit features from neurons and the interconnection features between them. Nowadays, on one hand, we already know the basic mechanism of one single neuron, and we have enriched recording data from neurophysiology to know the features and connectivity of small groups of neurons; on the other hand, experimental psychologists and behavioral neuroscientists managed to help us to understand the

systematic functional purpose of the visual system using tools such as visual illusions, psychophysics experiments, and functional magnetic resonance imaging (fMRI) to locate the neuronal activation in the brain. However, the connection from the macro world and the micro world are not so clear: most models proposed by neurophysiologists are only trying to explain the recording data without considering any functional role, and most models proposed by experimental psychologists and behavioral neuroscientists are only trying to explain the functional role without considering the neurophysiological plausibility.

Predictive coding is one of the theories that try to connect the macro and micro world. The idea behind predictive coding came long before the theory itself, from the efficient coding hypothesis. After the development of information theory in the end of 1940s, people wanted to explain the brain as a machine that reduces the amount of information and codes sensory input in the most efficient way. Under this influence, predictive coding is one implementation for this efficiency: the predicted response is inhibited by the feedback (prediction), and the feedforward signal only contains the difference between the incoming information and prediction. This method could dramatically reduce the information especially in a stable environment (most of the time of our daily life is in a stable environment) and it is a natural method to reduce the redundant sensory information.

The brain should have only one unique model. This model should be based on our unique nervous system, inherit the basic features of neurons, neuronal connections, and follow the known structure, and functional role of different parts of the nervous system. In this thesis, we

tried to implement the neuronal circuits for the modern understanding of the efficient coding: predictive coding.

In the introductory part of this thesis, I will review my current understanding about neuron, neocortex, and predictive coding.

In the main chapters, first, I will present the core of this thesis: a modeling work on how to use the correlated spike-time to generate selectivity in a non-selectivity excitatory network. This model is also a viable mechanism for predictive coding.

Then, I will present two empirical evidence on predictive coding: one is about the excitatory predictive feedback, the other is about the oscillations in predictive coding.

In the end, I will conclude my thesis. I will also comment on my current work and propose several possible future projects not only about the computational modeling, but also about the experiments.

Neuron

A huge tree that fills one's arms grew from the tiniest sprout; a tower of nine storeys rose from a heap of earth; a journey of a thousand miles commenced with a single step.

-Tao Te Ching, Laozi

Ramón y Cajal (1852 – 1934) is the first scientist that reported neurons as individual: he demonstrated experimentally that the relationship between nerve cells was not continuous, but contiguous, by studying the small, star-shaped cells of the molecular layer of the cerebellum of birds in 1888 (Cajal, 1888; López-Muñoz et al., 2006; De Carlos and Borrell, 2007). Then he successfully convinced the scientific community with the idea of the Neuron Doctrine: neurons are not connected in a meshwork, but discrete cells act as distinct units. Besides that, as a great histologist. Ramón y Cajal used Camillo Golgi's silver nitrate preparation method to stain the neurons and did a lot of drawing of neurons. In this thesis, the focus is on the neocortex, the mammalian (human) brain area involved in functions such as sensory perception, attention, motor control, language, and conscious thought (Lui et al., 2011). Thus, the information about neurons is mostly from neocortex.

From the end of the 19th century and the beginning of the 20th century, the pioneers of the field of neuroscience discovered several types of neurons and named them with their own names, such as the Golgi type I neuron, Golgi type II neuron (Camillo Golgi, 1843-1926), Purkinje neuron (Jan Evangelista Purkyně, 1787-1869), Lugaro neuron (Ernesto Lugaro, 1870-1940), Betz neuron (Vladimir Alekseyevich Betz, 1834-1894),

Martinotti neuron (Carlo Martinotti, 1859-1918). Since the microscope is the only method to observe, all these classifications of neurons were based on the physical properties of neurons, or specifically, the shape of neurons. After the discovery of the neurotransmitters in 1921 by Otto Loewi (1873-1961), neurons were classified based on the different types of neurotransmitters. Then, after the first account of being capable of recording electrical discharges in single nerve fiber in the neuronal system in 1928 by Edgar Adrian (1889-1977), neurons were classified based on their electrical features. The physical properties, the neurotransmitters, and the electrophysiological properties tell us what one neuron can do. Furthermore, the connections between these properties can tell us the possible functional roles for different types of neurons.

Physical properties

Shape

The most obvious feature of neurons under the microscope is their shape. However, since the descriptions of shape are subjective, there are many kinds of ways to describe a neuron's shape. In most neuro-anatomy books (Susan Standring, 2009; Watson, Kirkaldie, & Paxinos, 2010), the shapes of neurons are usually described based on the structural polarity: unipolar (or pseudounipolar, axon and dendrite from same process), bipolar (one axon, one dendrite) and multipolar (one axon, multiple dendrites). However, one simple way to describe the neuron shapes was

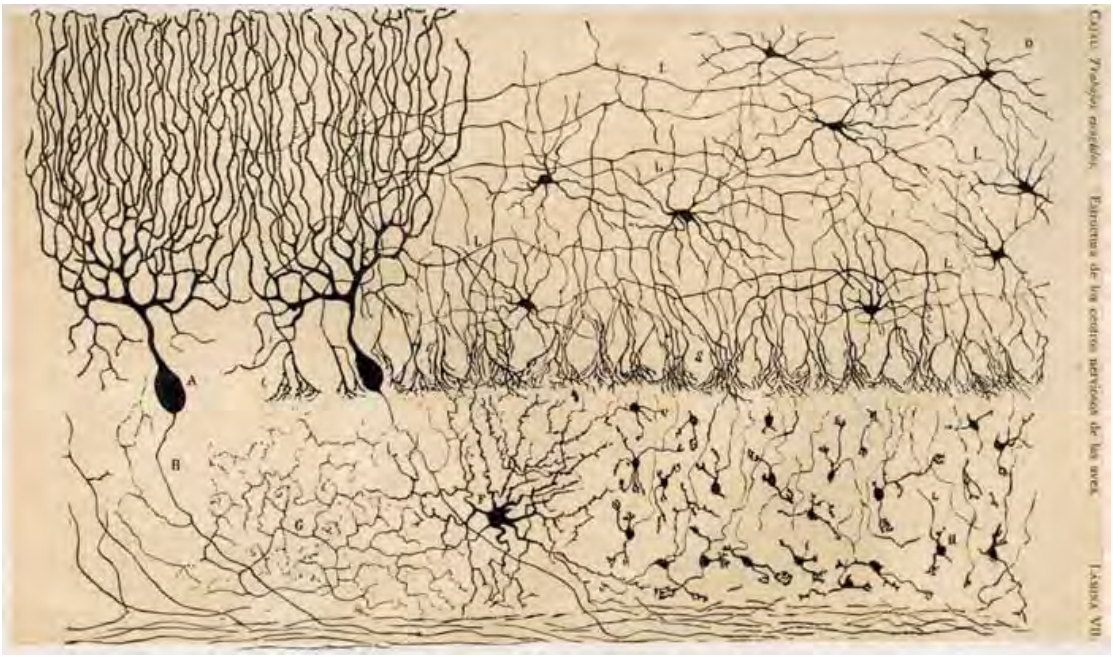


Figure 1-1 Drawing of different neurons by Ramón y Cajal. Five classes of neuronal populations of the cerebellum are in the picture: Purkinje, stellate, basket, Golgi and granule cells. Also note basket-cell axons terminating freely around the Purkinje cell bodies. A, Purkinje cell; D, stellate cell; F, Golgi cell; H, granule cell; S, basket cell axons.

proposed early (Sholl, 1956) with only two types of neurons: pyramidal and stellate neurons.

Pyramidal neurons have the cell body shaped like a pyramid, with a single axon and multiple dendrites(Abeles, 1991). The soma size of pyramidal neurons is about 20 micrometers (order of magnitude: 10^{-2}mm) (Larkman and Mason, 1990). The dendrite's diameter is from less than half to a few micrometers (order of magnitude: 10^{-3}mm). The dendrites could be divided into two types: basal and apical dendrites. The primary apical dendrite extends for several hundred micrometers before branching (order of magnitude: 10^{-1}mm). The linear distance from the basal end to the apical end of the dendritic tree is from two hundreds

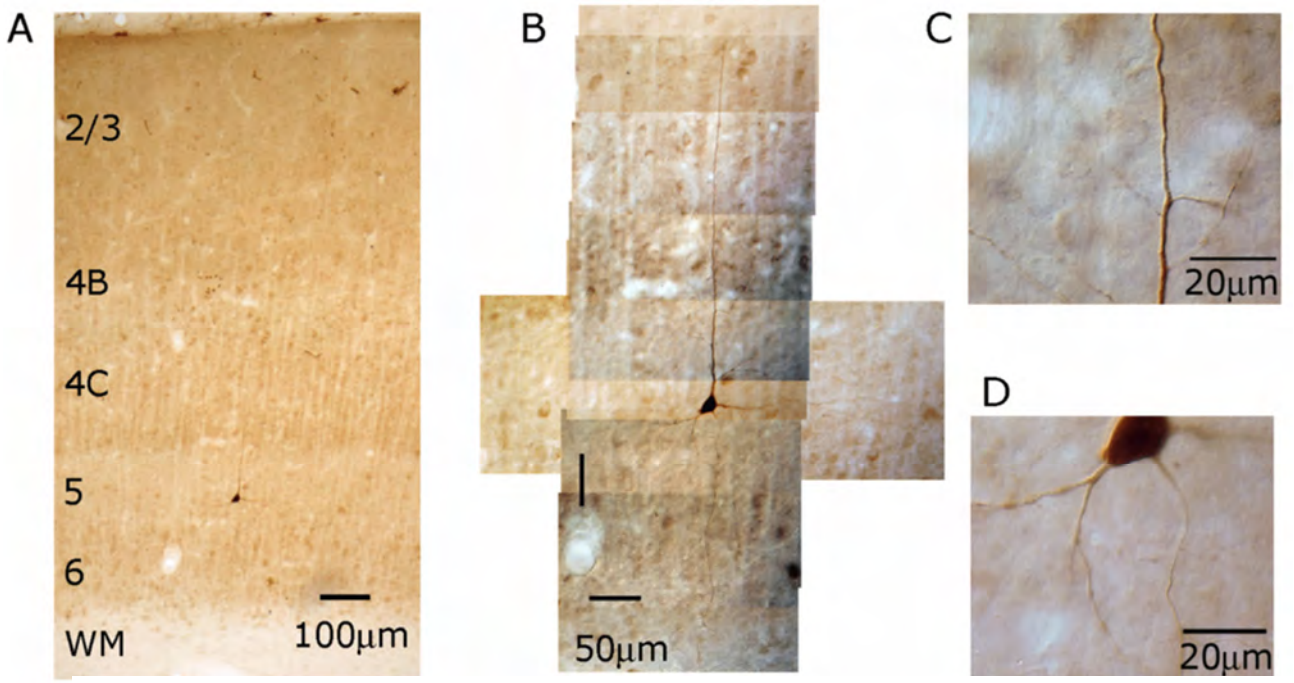


Figure 1-2 Morphology of layer 5 pyramidal neuron in macaque primary visual cortex. (Joshi, 2007)

to one thousand micrometers (order of magnitude: 10^{-1}mm). For the axon, the length could be tens of centimeters (order of magnitude: 10^{3}mm) (Spruston, 2009). In the cortex, at least more than half of the neurons in the cortex are pyramidal neurons.

Stellate neurons (see the Figure 1-3) have cell bodies shaped like a star, with a single axon and multiple dendrites extending from all aspects of the soma(Abeles, 1991). Stellate neurons have spherical or ovoidal cell bodies with the range from 9 micrometers to 14 micrometers (order of magnitude: 10^{-2} mm) (Wouterlood et al., 1984). The extending axon and dendrites create an axonal field with 100-150 micrometers (order of magnitude: 10^{-1} mm) and a dendritic field with 80-200 micrometers (order of magnitude: 10^{-1} mm) (Kisvarday et al., 1986). There are less stellate neurons than pyramidal neurons in the cortex (see Table 1-1).

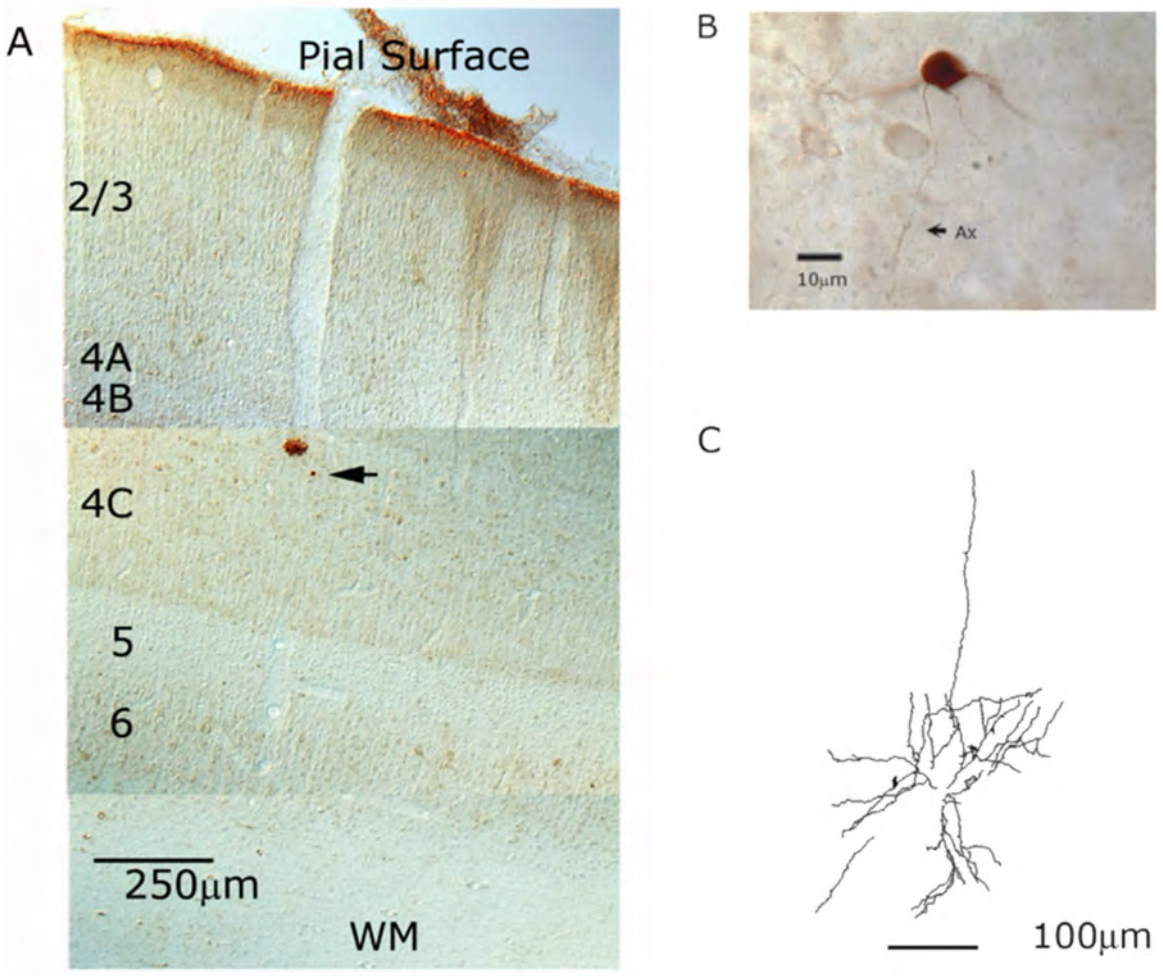


Figure 1-3 Spiny stellate neuron of layer 4B in macaque primary visual cortex. A and B are the morphology, C is the computer reconstruction of dendrite in horizontal view.

From these data, we could clearly see that it is not common for the stellate neuron to receive input from neurons in other areas directly. For this reason, we could also call the pyramidal neurons as principal neuron, and the stellate neuron as intrinsic neuron.

Animal	Region	Pyramidal	Stellate	Fusiform		Animal	Region	Pyramidal	Smooth stellate	Others
Cat	Visual	62%	34%	4%		Rabbit	Auditory	86.70%	9.50%	3.80%
	Somatosensory	63%	35%	2%		Rat	Visual II + III	87%		
	Motor	85%	10%	5%			IV	90%		
Monkey	Visual	52%	46%	2%			V	89%		
	Motor	74%	22%	4%			VIa	97%		
Human	Prefrontal	72%	26%	2%						

Table 1-1 The percentage of pyramidal neuron, stellate neuron in different regions in different animals. (Abeles, 1991)

Dendritic spine

One other physical property of the neurons discovered by Ramón y Cajal is the spine on the dendrite(Shepherd, 2004). There are two kinds of dendrites: one bears spines and another does not. Spines are filopodium, thin, stubby mushroom-shaped or cup-shaped with length of 0.5 - 2 micrometers (order of magnitude: 10^{-3} mm) (Hering et al., 2001) (see Figure 1-4), and are rarely found in lower organisms. In neocortex, both pyramidal neurons and stellate neurons can have the dendritic spines. However, most GABA-releasing interneurons do not have dendritic spines. In fact, depending on the number of spines the dendrites have, we can determine whether a stellate neuron is a GABA-releasing or glutamate-releasing neuron: the glutamate-releasing neurons have lots of dendritic spines, while the GABA-releasing neurons have few. Hence, we can divide the neurons into 3 classes: pyramidal neuron, spiny stellate neuron, and smooth stellate neuron. The pyramidal

neurons form about 70% of the neurons and smooth neurons from about 20% (Shepherd, 2004).

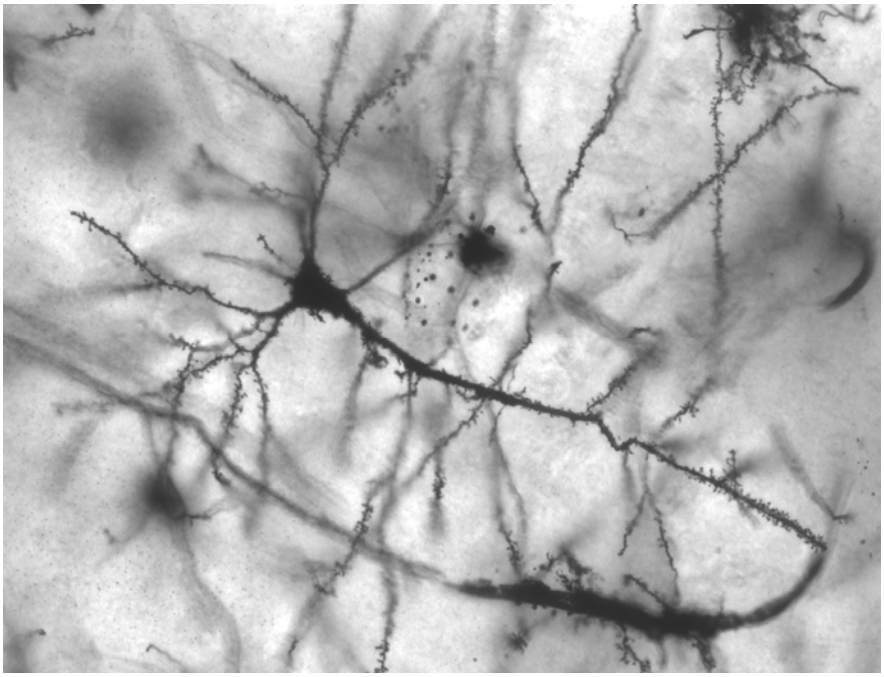


Figure 1-4 Hippocampus pyramidal neuron with dendritic spines. From MethoxyRoxy, Wikimedia Commons.

Neurotransmitters

As Ramón y Cajal suggested, the neurons are individuals, which means that the neurons are not sharing their electrical properties with other neurons (however, this is not true for the gap junctions between inhibitory neurons). Most neurons use chemical synapses to connect with each other. Most synapses connect axons to dendrites, but some also connect axons to soma. At the chemical synapse, the axon will release different neurotransmitters through the small gap of the synaptic cleft. The chemical synaptic cleft is about 20nm wide (order of magnitude: 10^{-}

$^{-6}$ mm) (Kendal et al., 2000). With such a short distance, the complicated chemical signal transmission process could be finished in less than 1 ms (usually 0.5 ms). Depending on the neurotransmitters one neuron releases, we can classify the neurons into two types: excitatory neurons and inhibitory neurons. These two kinds of neurotransmitters have very different effect on the reception neurons: the glutamate-releasing (or similar chemical material such as Acetylcholine, Catecholamines, Serotonin, Histamine) neuron makes the reception neuron fire more and the GABA-releasing (or similar chemical material such as GABOB, Proline) neuron makes the reception neuron fire less (Kendal et al., 2000; Shepherd, 2004). These two types of neurons are the keys to understand the dynamics of neuronal process in the brain.

Excitatory neurons

Based on the shape features described before, we know that there are two types of excitatory neurons: pyramidal neurons and spiny stellate neurons. There are many more pyramidal neurons than spiny stellate neurons: about 75% of the neurons in the cortex are pyramidal neurons, while only about 10% of the neurons are spiny stellate neurons (Abeles, 1991). Furthermore, these spiny stellate neurons are only in the cortical sensory areas. In the non-sensory areas, there are few spiny stellate neurons; in some animals, there are no spiny stellate neurons outside sensory areas (Peters and Kara, 1985a, 1985b).

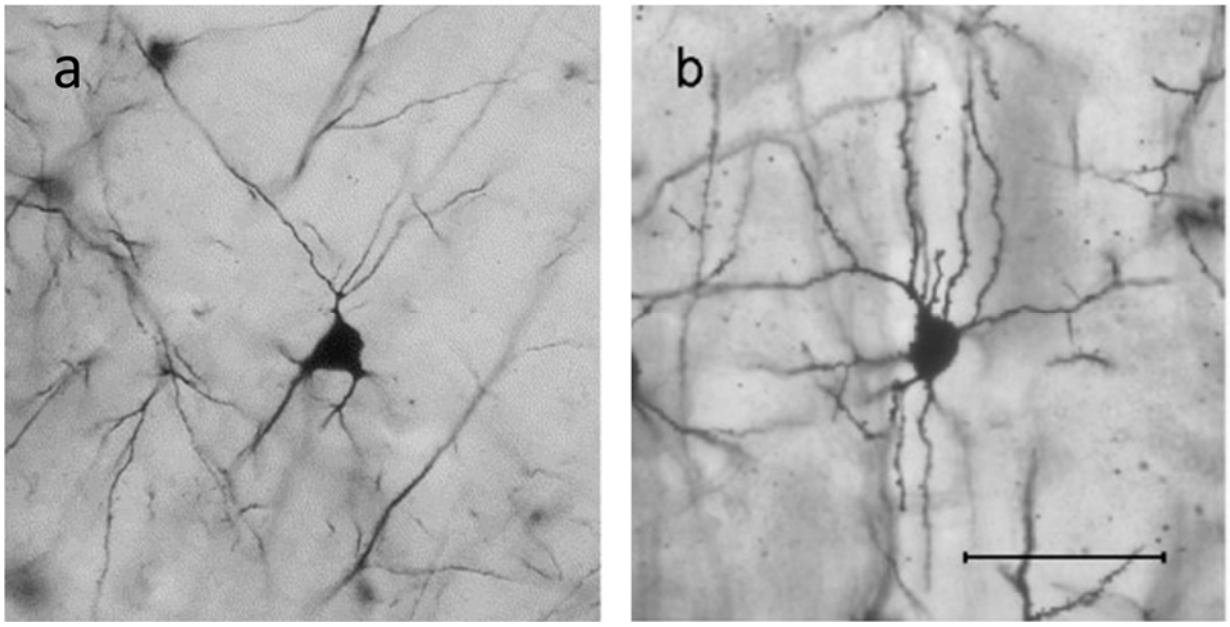


Figure 1-5 Smooth stellate neuron and spiny stellate neuron. Photo a is a smooth basket neuron in human prefrontal cortex(Benes and Berretta, 2001), photo b is a monkey layer IV spiny stellate cell(Churchill et al., 2004).

In the connections between excitatory neurons (including pyramidal and spiny stellate neuron) and pyramidal neurons, most excitatory synapses are made on the spines of post-synaptic neurons (65%-85%) while only most of excitatory synapses on spiny stellate neuron are on the shafts of dendrites (~60%). Excitatory synapses never land on the soma (Shepherd, 2004).

Excitatory synapses generate excitatory post synaptic potential (EPSP). For one single neuron, the effect of the EPSP is very small. When there is a spike, and recording in the neuron body, one single spike only provides a 0.4-1 mV (Mason et al., 1991; Markram and Tsodyks, 1996) increase in the voltage between pyramidal neurons, and an about 1.5mV increase between spiny stellate neurons. The variability in the voltage gain

between pyramidal neurons are huge (from 0.05mV-2.08mV) (Mason et al., 1991).

Inhibitory neurons

Even though we can classify the inhibitory neurons into more than 10 categories, the basic shape according to the shape classification method mentioned before, the inhibitory neurons are all in the same category: stellate neurons. However, at the same time, depended on the expressed genes, all the inhibitory neurons can be classified into 3 main classes: Htr3a, Pvalb and Sst (See Table 1-2).

Table 1-2 The three classes of inhibitory neurons and the inhibitory neurons that falls into the category. (Harris and Shepherd, 2015)

Top-level class:	Htr3a		Pvalb		Sst	
Subclass:	Vip	Neurogliaform	Basket	Chandelier	Martinotti	L4 Sst
Local outputs:	Descending axon, inhibiting Sst and Pvalb	Nonsynaptic GABA release	Inhibiting ECs (soma), other Pvalb	Inhibiting axon initial segment of ECs	Inhibiting Pvalb, EC dendrites including tufts	Inhibiting L4 Pvalb
Local inputs:	Excited by ECs	Excited by ECs	Excited by ECs, inhibited by Pvalb, Sst, Vip	Excited by ECs	Excited by ECs, inhibited by Vip	Excited by ECs
Long-range input:	Higher order cortex	?	Thalamus, lower order cortex	?	?	?
Intrinsic physiology:	Irregular spiking	Late spiking	Fast spiking	Fast spiking	Low threshold spiking	Intermediate fast/low-threshold pattern
In vivo activity:	Driven by behavior	?	Dense code, weakly tuned	?	Modulated by motor activity; wide visual receptive fields	?

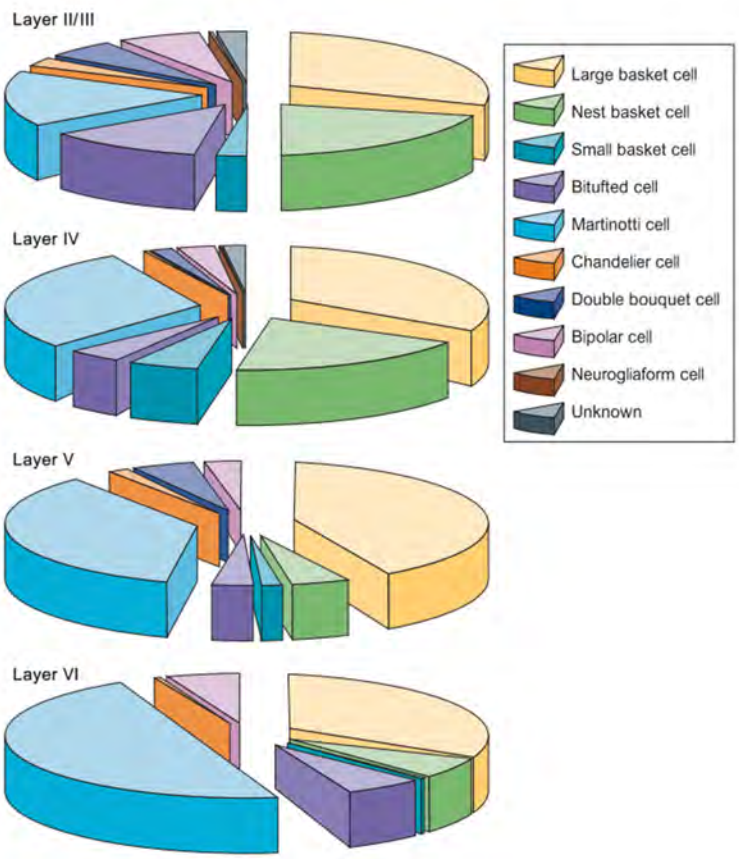


Figure 1-6 The percentage of different types of inhibitory neurons in different cortical layers. (Markram et al., 2004)

From Figure 1-6 we can see that, most of the inhibitory neurons (about 50% of all inhibitory neurons) are basket neurons (Markram et al., 2004). Basket neurons have a shape of basket with extensive axons to form lateral connections with 300-500 other neurons (most of them are pyramidal and spiny stellate neurons) with 10 synapses on average(Shepherd, 2004).

Depending on different neuron types, the connection place between the inhibitory neuron and the reception neuron is different. For example, basket neurons connect their axon to the dendritic shaft and spines,

while the chandelier neurons connect their axon to the axon (Shepherd, 2004). One other type of inhibitory neuron that was studied a lot is the Martinotti neuron. It was found that Martinotti neurons are linked to the cortical dampening mechanism by sending inhibitory signals to the surrounding neurons (Silberberg and Markram, 2007).

Inhibitory synapses generate an inhibitory postsynaptic potential (IPSP). For one single neuron, compared to the EPSP, the effect of the IPSP is big. When there is a spike, and recording in the neuron body, one single spike provides a 10 mV decrease in the voltage in pyramidal neurons. IPSPs reach peaks at about 20-30 ms and have a duration of 200-300 ms (Shepherd, 2004). The inhibitory synapses land on the soma or the proximal dendrites, this could be one reason that inhibitory synapses have a bigger effect than excitatory synapse.

Table 1-3 The order of magnitude of different parts of the neurons and neocortex.

	Property Name	Order of magnitude
Pyramidal neurons	Soma size	10^{-2} mm
	Dendrite's diameter	10^{-3} mm
	Length of primary apical dendrite	10^{-1} mm
	Length from the basal end to the apical end	10^{-1} mm
	Length of axon	10^{3} mm
	Dendritic spine	10^{-3} mm
Stellate neurons	Body size	10^{-2} mm
	Axonal field	10^{-1} mm
	Dendritic field	10^{-1} mm
Neurotransmitters	Chemical synaptic cleft	10^{-6} mm
Neocortex	Thickness	10^{0} mm
	Distance between areas	10^{1} mm

Electrophysiology

One other important feature about the neuron is its electrophysiological properties. Since communication between neurons is based on spikes, it is important to know how the neuron spikes, the effect after the neuron's firing (AHP or ADP), and is there any difference between different neuron groups' effects and if there is, what is the possible reason for that.

In this part, I took advantage of the data gained from 64 studies to investigate the electrophysiological properties of neurons. I created a database of studies including different neurons from different locations of neocortex based on a previous electrophysiological database (Tripathy et al., 2015). I show the statistical values of the properties gained from the data. The statistical value contains the medium value and the standard derivation.

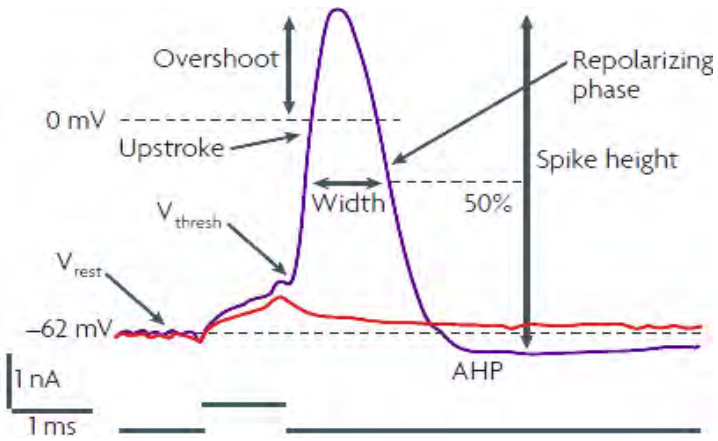


Figure 1-7 Different electrophysiological features of the action potential of a neuron.

Since information exchange in the brain takes advantage of the spike, or action potential, all of the electrophysiological properties are about that. For the action potential, as Figure 1-7 shows, there are different parts:

- Resting membrane potential
- Spike threshold
- Spike amplitude (spike height)
- Spike width
- Input resistance
- Membrane time constant
- Firing frequency
- Fi slope
- AHP amplitude
- AHP duration
- Adaptation ratio

Depending on their values in different types of neurons, we divide the properties into two categories: with similar values across different types of neurons and with different values across different types of neurons. Here, I will show the definition of these properties, the regular measure method in electrophysiology experiments and statistical values across studies. Then, I will speculate the possible reasons for the patterns of the values.

Properties with similar values across different neurons

Resting membrane potential

Resting membrane potential is the membrane potential in the "balanced state" or "resting state" of a neuron. The membrane potential is usually recorded using the patch clamp technique. The resting membrane potential is usually defined as the membrane potential of a neuron going for a long period of time without changing significantly. This "balanced state" could be described using the Goldman equation (Koch, 1998):

$$E_m = \frac{RT}{F} \ln \left(\frac{\sum_i^N P_{M_i^+} [M_i^+]_{\text{out}} + \sum_j^M P_{A_j^-} [A_j^-]_{\text{in}}}{\sum_i^N P_{M_i^+} [M_i^+]_{\text{in}} + \sum_j^M P_{A_j^-} [A_j^-]_{\text{out}}} \right)$$

Where the E_m is the membrane potential, P_{ion} is the permeability for the ion. $[ion]_{out}$ and the $[ion]_{in}$ is the extracellular and intracellular concentration of that ion, T is the temperature and R and F are constant. This equation describes that the resting membrane potential has relationship with the ion flow. The stable values in Figure 1-8 indicate that the ion channels of these neurons should be similar and the inside and outside environment of the neurons should be similar.

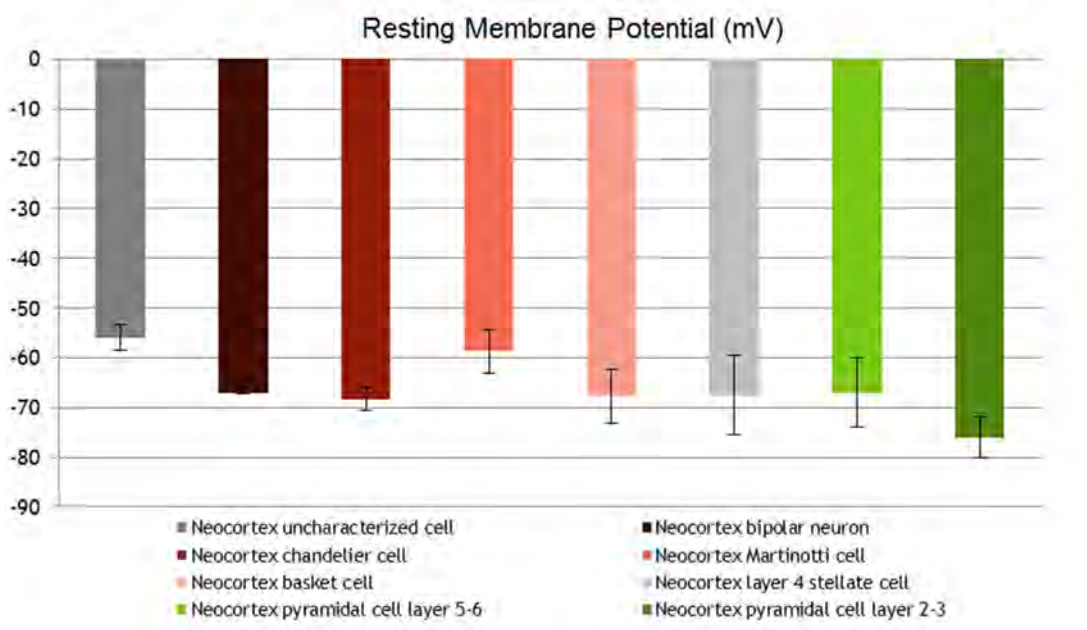


Figure 1-8 Resting membrane potential in different types of neurons

Spike threshold

Spike threshold is the voltage needed to initiate the action potential. It is usually measured by using the sudden rising slope of membrane voltage. We can see from Figure 1-9 that the spike threshold is constant in different types of neurons, which indicates that the biophysical requirements for action potential are similar in different kinds of neurons (they may use the same Hodgkin-Huxley theory for spike generation(Hodgkin and Huxley, 1952)). However, some argue that this similarity is not true for the spike threshold in *vivo*. Studies showed that the spike threshold of cortical neurons has a larger variability and could be adapted over time. The variabilities are different in different kinds of neuron. They propose that the rate of rise of pre-spike membrane

potential (Azouz and Gray, 2000; Henze and Buzsáki, 2001; Fontaine et al., 2014) and the recent history of spikes (Henze and Buzsáki, 2001) correlates with the spike threshold. Evidence suggested that this mechanism increases the sensitivity to simultaneous synaptic inputs and functions as a coincidence detector (Azouz and Gray, 2000; Howard and Rubel, 2010). On the other hand, other authors suggested that this threshold variability observed *in vivo* reflects only measurement artifacts (Yu et al., 2008). Other researchers have suggested a lower spike threshold for the basket neurons (Buzsáki and Wang, 2012), however, this difference is not so obvious in an inter-studies point of view. (Figure 1-9)

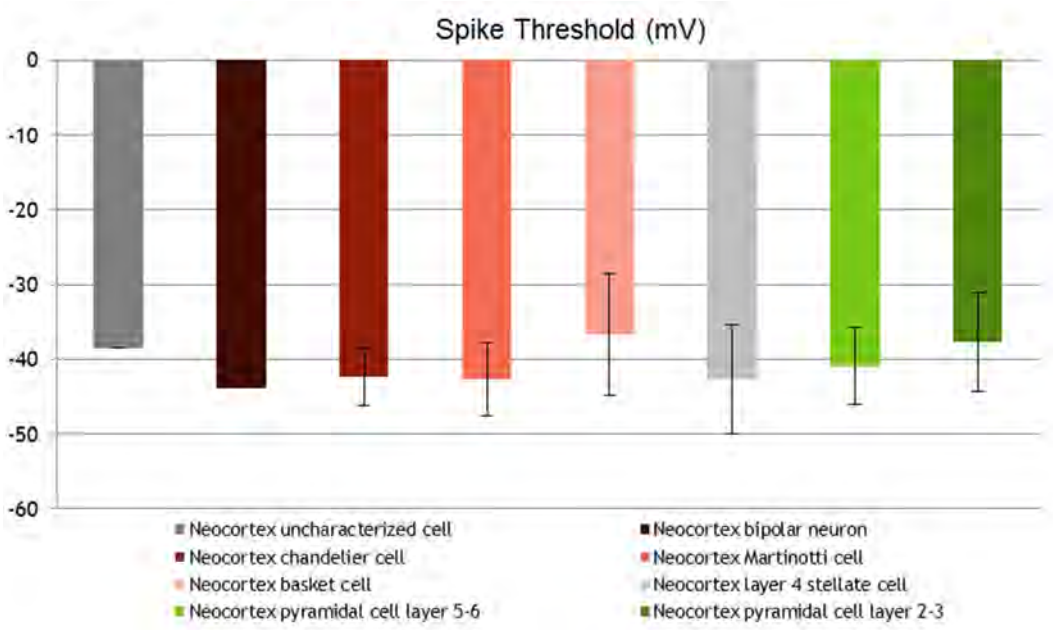


Figure 1-9 Spike threshold in different types of neurons

Spike amplitude

Spike amplitude (spike height) is the height of the action potential. It is usually measured by calculating the difference between the peak of the action potential and the threshold of the action potential (or the AHP) using the first spike of the spike train.

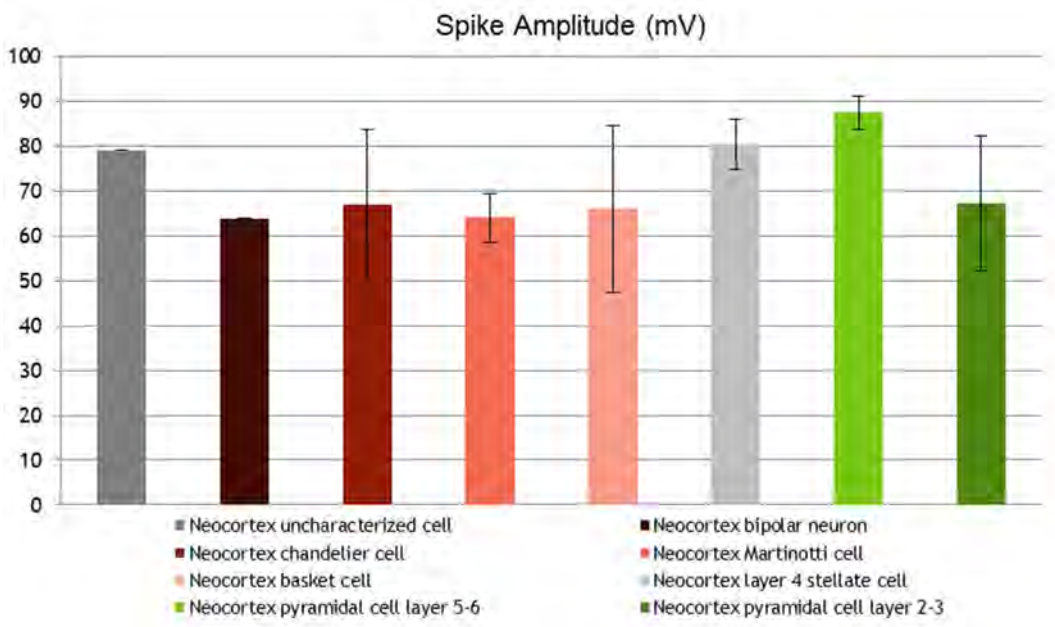


Figure 1-10 Spike amplitude in different types of neurons

Properties with different values across different neurons

Input resistance and membrane time constant

Input resistance is calculated using Ohm's Law: $R=V/I$, where R is the resistance and V is the voltage increase, and I is the input current in the depolarization stage of the action potential. Input resistance is usually measured at steady-state voltage response to current injection.

However, for most membranes (e.g. soma), the voltage has a non-linear relationship with the input current (since the neurons are more like a RC circuit rather than simply the resistance), we can use the time constant to measure the voltage-input relationship more accurately (the time constant is a parameter of the exponential voltage). From the collected data, we can see that pyramidal neurons and basket neurons have relatively small input resistance/time constant. On the other hand, we can see that the inhibitory neurons express very different voltage-related properties.

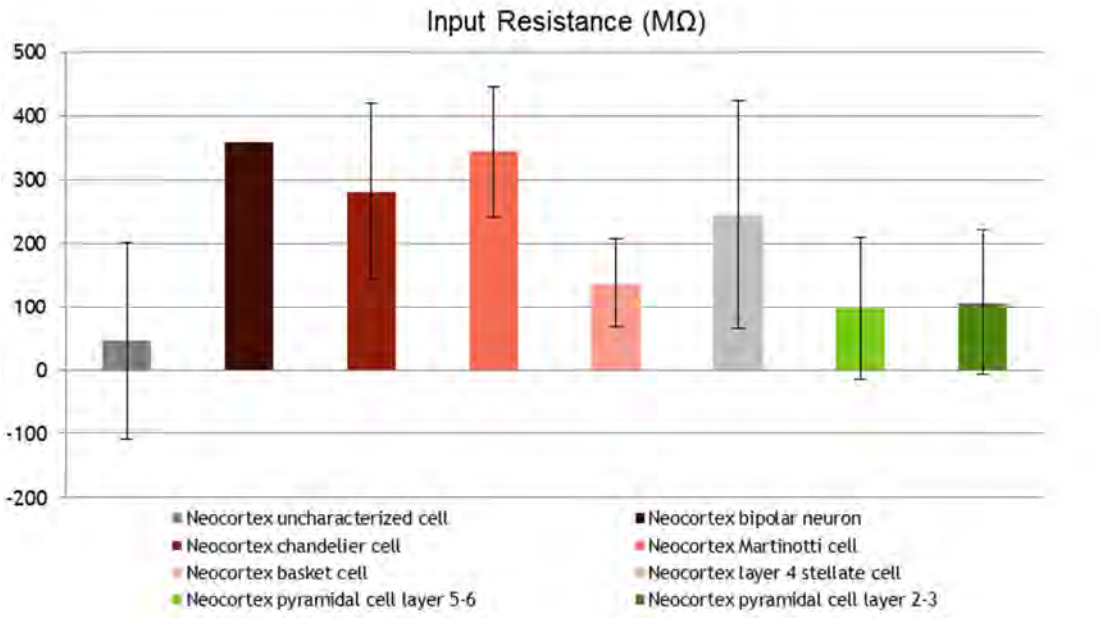


Figure 1-11 Input resistance in different types of neurons

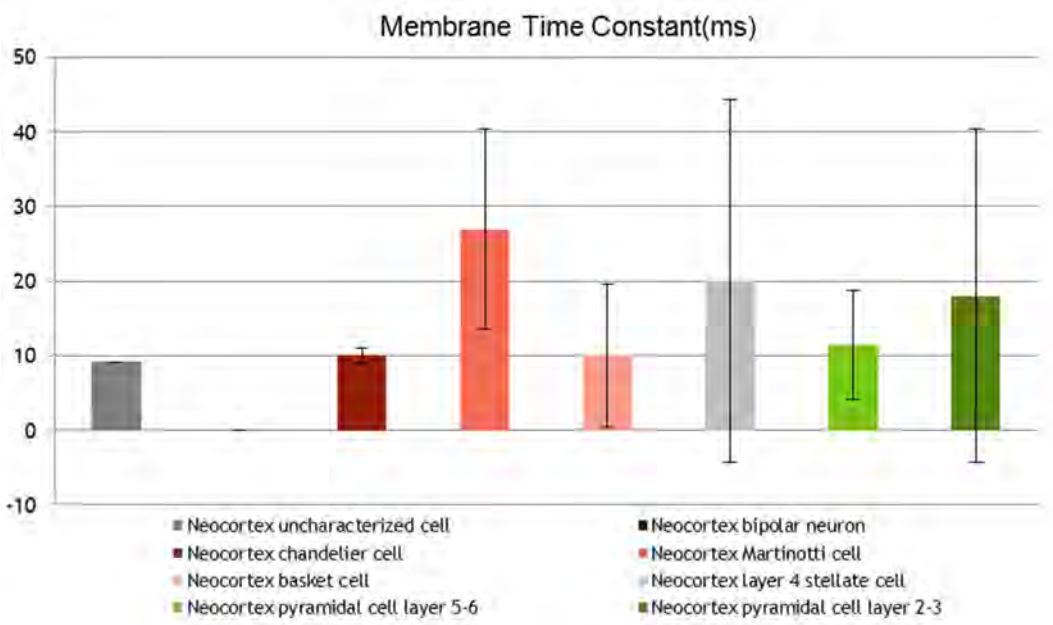


Figure 1-12 Membrane time constant in different types of neurons

Spike width

Spike width is most often measured as the width at half-maximal spike amplitude. Spike width is one of the most obvious electrophysiological properties that difference between GABA-releasing neurons and glutamatergic pyramidal neurons: GABA-releasing neurons have narrower spikes than glutamate-releasing neurons(Bean, 2007). From the Figure 1-13, we can see it is true for the basket neurons and chandelier neurons; however, this is not true for the Martinotti neurons. Thus, we can distinguish the different types of inhibitory neurons and indicate a different role for Martinotti neurons.

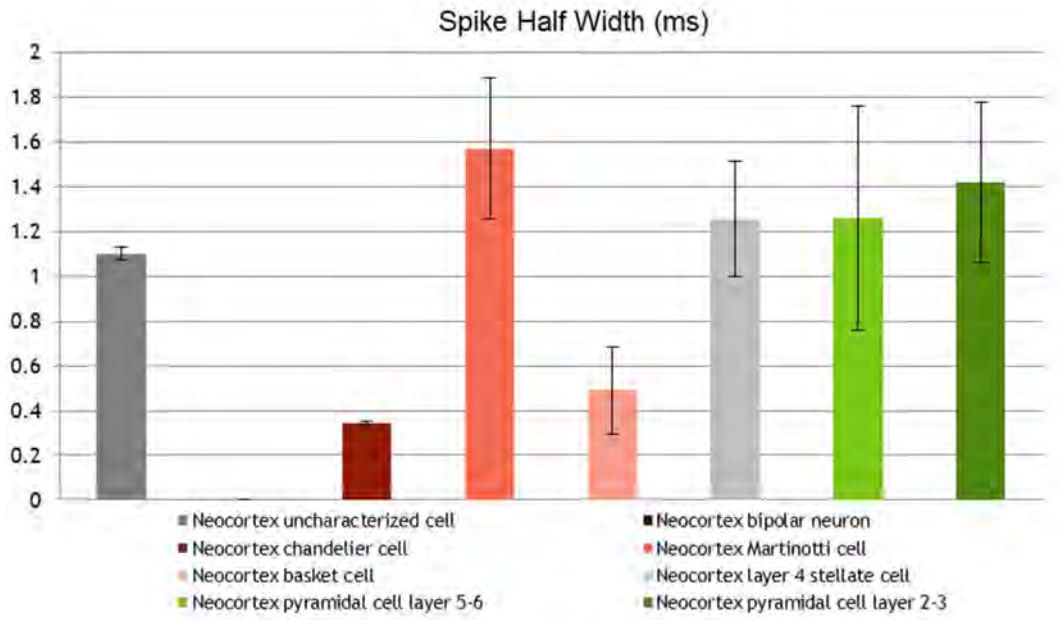


Figure 1-13 Spike half width in different types of neurons

Firing frequency and Fi Slope

This is the firing frequency of different types of neurons by injecting different amounts of current. We can see from Figure 1-14 that basket neurons fire much faster than all other kinds of neurons (this should be the reason that inhibitory interneurons are often called fast-spiking neurons). However, as we can see, Martinotti neurons have a rather low firing rate. This information also provides evidence that Martinotti neurons are very different from the fast-spiking inhibitory neurons as we understand. Because of this difference in firing frequency, Martinotti neurons have been linked to theta-band oscillations, while basket were neurons linked to gamma-band oscillation (Fanselow et al., 2008; Buzsáki and Wang, 2012).

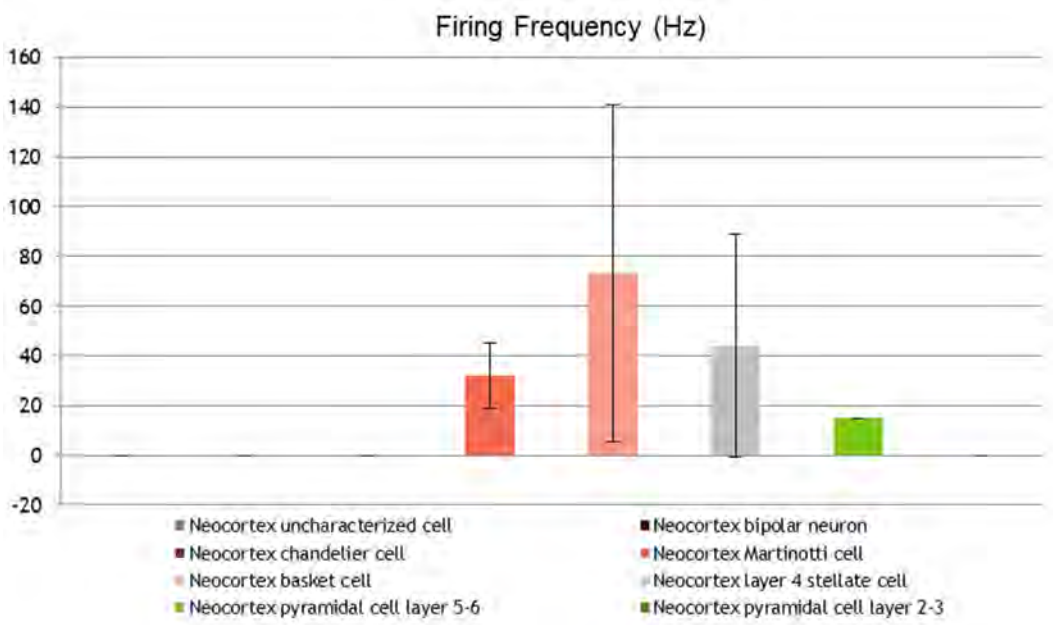


Figure 1-14 Firing frequency in different types of neurons. We only obtained the firing frequency for neocortex Martinotti cell, basket cell, layer 4 stellate cell and pyramidal cell in layer 5-6.

On the other hand, the Fi slope normalizes the input current and suggests a linear frequency-current relationship (which is obviously not true). But from the Figure 1-15, we still can find a similar pattern as the firing frequency. Note that the Martinotti neurons do not have fast-spiking properties, but rather behave like an excitatory neuron.

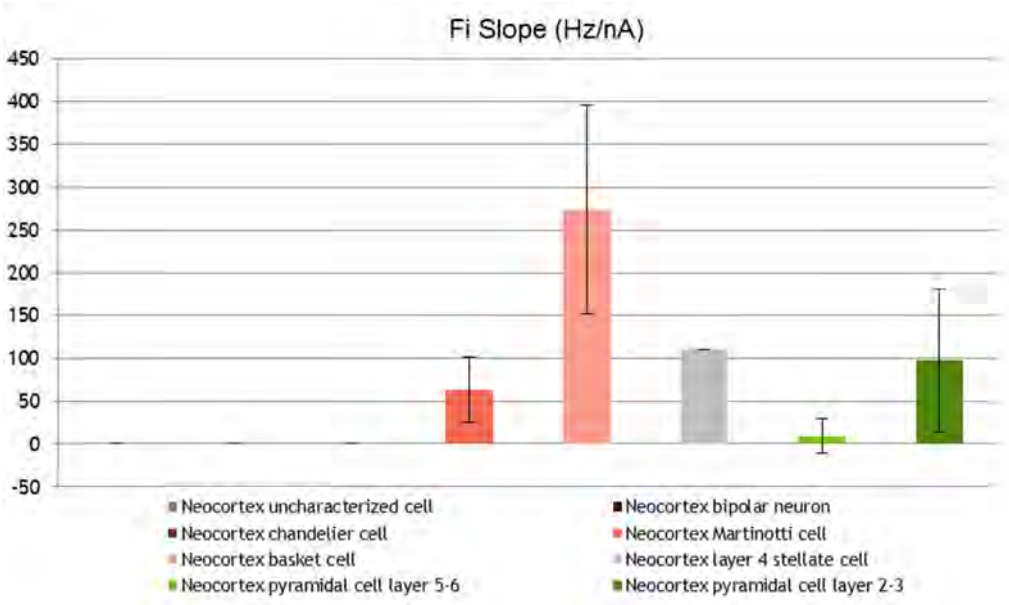


Figure 1-15 Fi Slope in different types of neurons

Afterhyperpolarization (AHP)

AHP is the hyperpolarized membrane potential after a neuron's action potential. It falls below the normal resting potential. It is also possible for depolarization to occur after a neuron's action potential (ADP) which usually is linked to the bursting neurons. For the pyramidal neurons, after the action potential, there is a fast AHP followed by an ADP, then there will be a slow AHP. The fast AHP is short (about 1 ms), the ADP is longer but in the same order of magnitude (about 5 ms), while a slow AHP is much longer (150-200 ms). From the Figure 1-16, we can see that for different types of neurons, the AHP amplitudes are different.

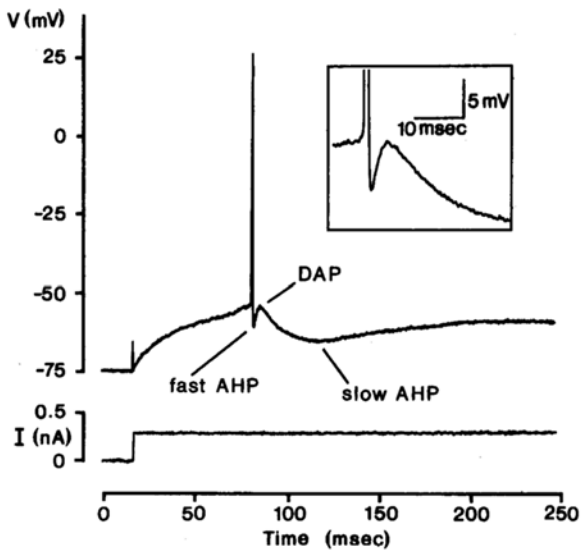


Figure 1-16 Action potential and after potentials in pyramidal neurons. (Mason and Larkman, 1990)

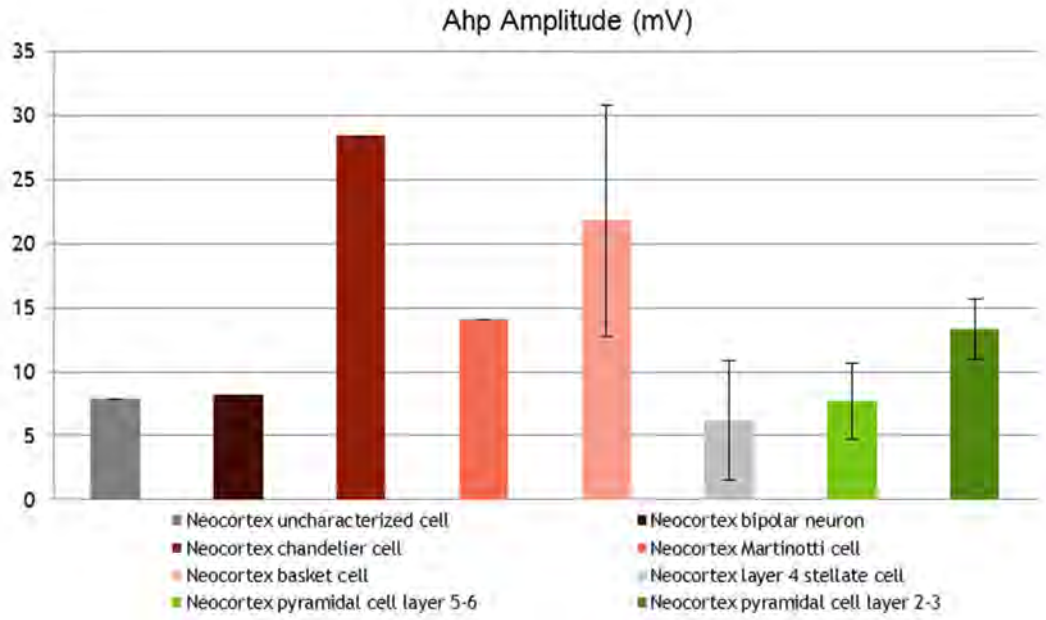


Figure 1-17 Ahp amplitude in different types of neurons

Adaptation ratio

Adaptation ratio is the ratio of durations between early and late AP inter-spike intervals in an AP train. The Figure 1-18 shows that inhibitory neurons have less adaptation than excitatory neurons except Martinotti neurons. This suggests that basket neurons can keep firing at a high frequency.

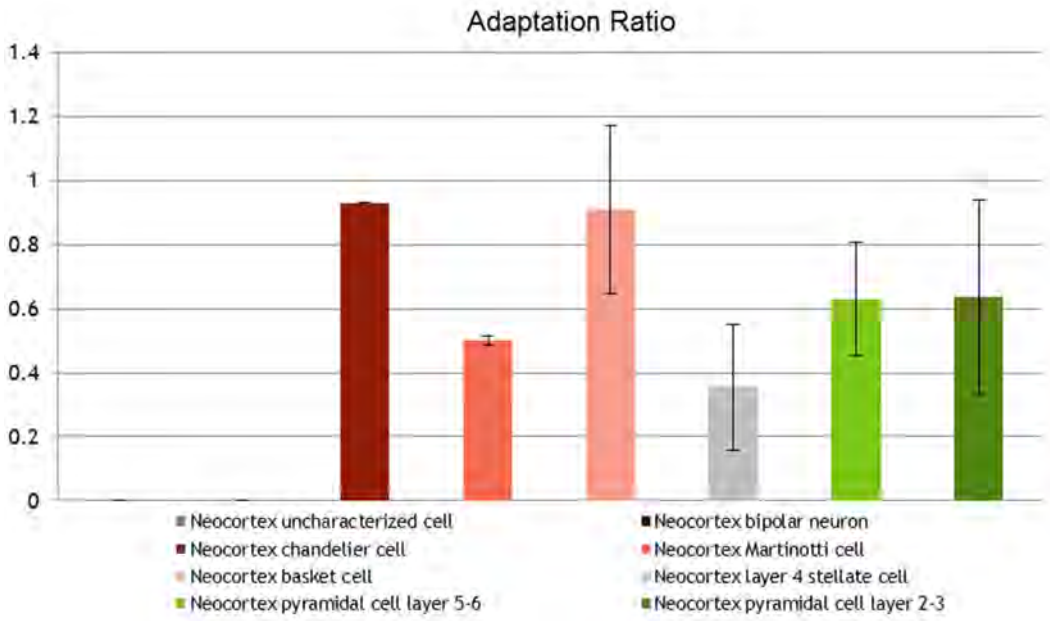


Figure 1-18 Adaptation ration in different types of neurons

Neocortex

If you do not make headway understanding a complex system, study its structure and knowledge of the function will follow automatically

- Francis Crick

Neocortex is the highest center of the brain. We understand it as the functional center of visual perception, auditory perception, motor controlling, reasoning, language and conscious thought. We can easily distinguish this part of the brain from the cerebellum, hippocampus, superior colliculus and some other areas. We consider various mammals, including rat, rabbit, monkey and human beings, have similar cortex in their brain.

However, as pointed out by Douglas and Martin (2007), the only property that defines neocortex is the "six layer" structure, in which the number of the layers could be subjective: depending on the areas and the histological stains used to reveal the layers, the number varies. Thus, the concept of neocortex itself is rather vague (Douglas and Martin, 2007).

We can divide approaches investigating the neocortex in history (and even nowadays) into two kinds: one is trying to find out the features of neocortical areas for different cognitive functions; the other is trying to find out the basic circuit for the neocortex. Since we still know little about the neocortex, we could say that both approaches are still ongoing. Scientists have tried a lot to understand the neocortex by looking at the human brain, even though the progress is slow.

Structure of neocortex

Marco-structure of neocortex

In the investigation of the functional roles of the neocortex, one well known metaphor for the functional organization of the neocortex is "Swiss Army Knife": the neocortex has a series of special-purpose modules as the Swiss Army Knife, such as the modules for vision, audio, language and etc. (Douglas and Martin, 2007).

Ironically, this idea first came from Franz Joseph Gall (1758-1828), the father of phrenology (1796, the pseudoscience claiming the size and shape of people's head is linked to their characters and abilities) (Mountcastle, 1995). More than half a century after the creation of phrenology, by studying the brains of aphasic patients (persons with speech and language disorders resulting from brain injuries), Paul Broca (1824-1880) discovered the area specifically devoted to speech processing (~1861). Then, Vladimir Betz (1834-1894) discovered the motor area (~1874). He also first divided the brain into eight regions, including the anterior central convolution, the arciform convolution, the hippocampus, the third frontal convolution, the lobules paracentralis, the gyrus lingualis, lobules extremus and the ventral extremity of the polus temporalis. In 1881, Hermann Munk (1839–1912) won the debate with David Ferrier (1843-1928) and confirmed that the visual area is in the occipital lobe (Glickstein, 1988).

Later, inspired by the influential evolutionary theory (~1862) of Herbert Spence (1820-1903), Hughlings Jackson (1835-1911) proposed his idea of a hierarchical brain (~1882). He stated that the brain is a sensorimotor machine and different brain regions represent the different hierarchical

levels, e.g. Anterior spinal horns and homologous cranial motor nerve nuclei represent the lowest motor level, the motor cortex and the basal ganglia represent the middle motor level and the premotor frontal cortex represents the highest motor level. Furthermore, he thought that the relationship between different hierarchical regions is that higher areas inhibit lower areas (York and Steinberg, 2006) which is consist with the modern theory of predictive coding (Rao and Ballard, 1999). He stated:

The higher nervous arrangements evolved out of the lower keep down those lower, just as a government evolved out of a nation controls as well as directs that nation. If this be the process of evolution, then the reverse process of dissolution is not only a "taking off" of the higher, but is at the very same time a "letting go" of the lower.(Jackson, 1882)

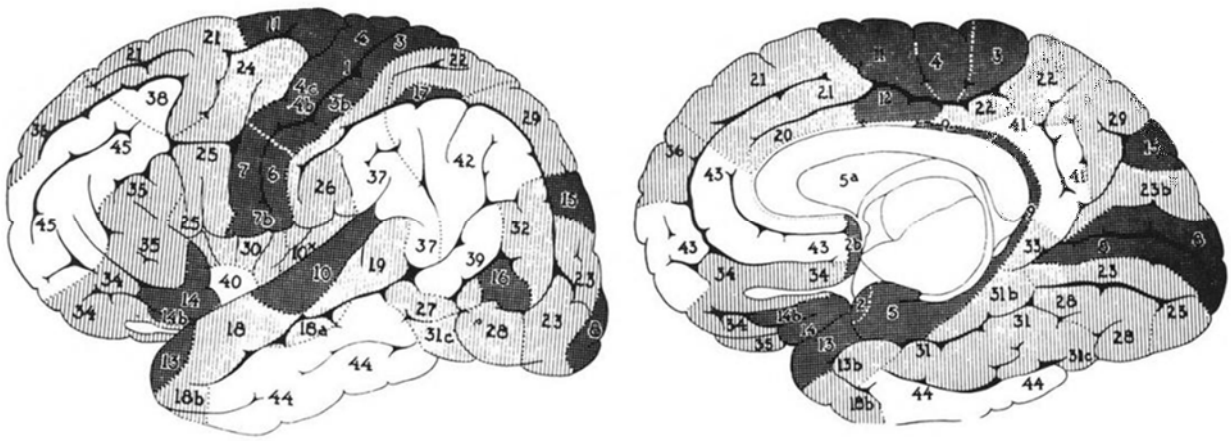


Figure 1-19The areas charted by Paul Flechsig. The shaded ones are the "primordial" areas and the white ones are the "association" areas of the cerebral cortex.(Flechsig, 1920)

Afterwards, by examining the distribution of myelination of the fibers in the white matter immediately subjacent to the cortex (they call it

myeloarchitecture), Paul Flechsig (1847-1929) divided the brain into 40 areas (~1896) and kept modifying this number ever since.

After entering the 20th century, Alfred Walter Campbell (1868-1937) and Korbinian Brodmann (1868-1918) followed Flechsig's work and continued to make a contribution to the lamination histology and the development of their own brain maps with their own observations and different naming styles.

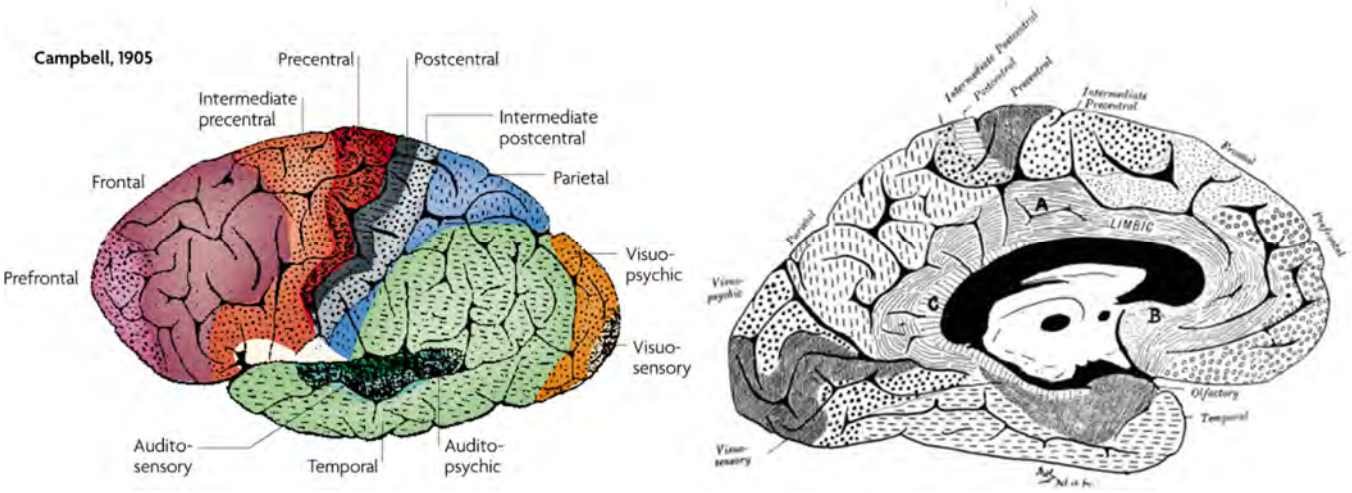


Figure 1-20 The areas charted by Alfred Walter Campbell.(Campbell, 1905)

Campbell showed a brain map with 14 areas in the cortex based on a 41 years old man (Campbell, 1905). Based on his studies (mainly autopsy of the patients with functional disability in rainhill mental hospital), he described the brain with a surprising modern information. He connected the functional role with different brain areas. For example, he described the vision areas (he even parted the visual areas into visuo-sensory and visuo-psychic) and auditory areas. He defined the precentral area as the motor area and the postcentral area as the sensory area(Campbell, 1905).

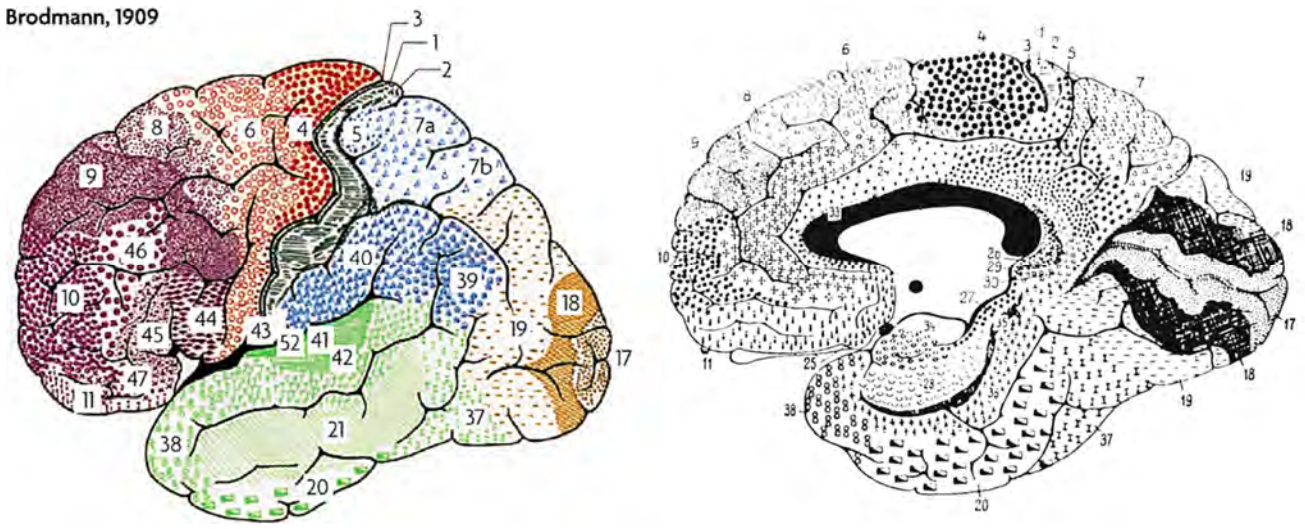


Figure 1-21 The areas charted by Korbinian Brodmann. (Brodmann, 1909)

Brodmann classified a more detailed brain map with 52 areas with 2 areas only existing in primates based on his continuous work on different species (including human, guenon, marmoset, lemurs, flying fox, Cercoleptes caudivolvulus, rabbit, ground squirrel and Erinaceus europaeus). He defined 11 homologous regions in man and other mammals including postcentral region, precentral region, frontal region, insular region, parietal region, temporal region, occipital region, cingulate region, retrosplenial region, hippocampal region and olfactory region. Since most textbooks copied either Campbell's or Brodmann's brain map, their maps became the standard brain maps that we are still using nowadays.

However, there were strong opposition opinions on these kinds of area classifications. For example, as Bailey and von Bonin pointed out, because of the sudden death of Brodmann, he did not finish his detailed

description of the areas indicated on his map of the human brain (he did that for the cercopithecus), and it is strange that the scientific world has accepted Brodmann's brain map, for which no direct proof had ever been given. They also made photos over 300 sites in the cerebral cortex and they found out in most cases, they could not correlate the photos to the cortical positions and they made the conclusion that most brain areas cannot be distinguished by pure cytoarchitectonic criteria (Bailey and Bonin, 1951). Since Oskar Vogt (1870-1959) and his followers classified the brain into more than 100 areas, Bailey and von Bonin pointed out these subtle distinctions between different areas are "hair-splitting" and have little influence (Bailey and Bonin, 1951; Jellison et al., 2004). Furthermore, they suffered the same problem: no subject-wise variations were taken into account in their brain maps.

Despite the criticism, we still use Brodmann's map nowadays. Thus, we can consider the 1909 Brodmann's areas as the state-of-the-arts of our knowledge of different brain areas. Brodmann's areas gained unexpected popularity because of the development of the functional and structural neuroimaging technique. These kinds of reports were heavily relayed on these brain areas classification and Brodmann's map was thus built into the processing softwares (Jellison et al., 2004). However, because of the criticism, a "Brodmann area" does not need to link to any functional meaning that Brodmann himself described or even imagined (Passingham and Wise, 2012).

A more detailed human brain map (1925) with structure was created by Constantin von Economo (1876-1931). He divided the cortex into seven lobes (Lobi) with further subdivisions (Regiones and Areae), the lobes (Economo and Koskinas, 1925) are: Lobus frontalis (with 35 areas), Lobus limbicus superior (with 13 areas), Lobus insulae (with 6 areas), Lobus

parietalis (with 18 areas), Lobus occipitalis (with 7 areas), Lobus temporalis (with 14 areas) and Lobus limbicus inferior/Lobus hippocampi (with 14 areas).

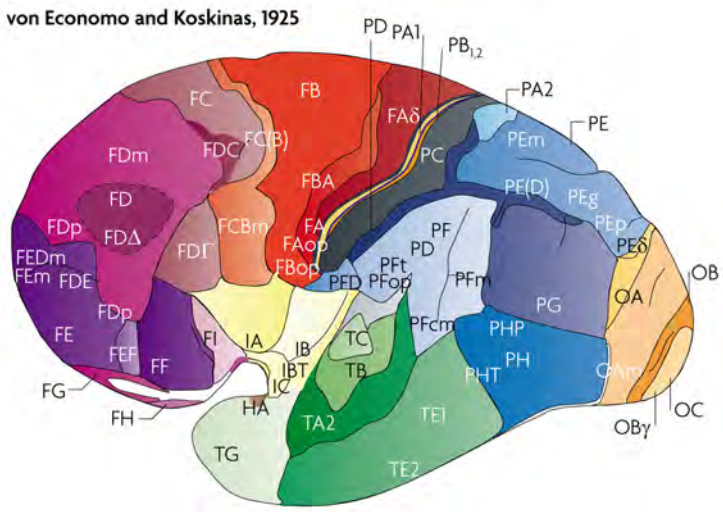


Figure 1-22 The areas charted by von Economo and Koskinas. (Economo and Koskinas, 1925)

Modern monkey brain maps with letter naming system were developed by Von Bonin and Bailey based on architectonic subdivisions, they divided the brain areas not only according to the cytoarchitectonic criteria, but also according to the study of monkey brain architecture (Bonin and Bailey, 1947). Another widely used brain map is from Van Essen's group. In 1980s, they were not satisfied with unclear borders with the Brodmann's map, since Brodmann's map was drawn on the three-dimensional brain. Thus, they developed a technique to unfold the cerebral cortex (Van Essen and Maunsell, 1980). They applied such technique on monkeys and obtained an accurate two dimensional map. They published the influential paper on brain area structure in 1991 (Felleman and Van Essen, 1991). The other maps by their group and other groups were also included in the brain image database such as

the Scalable Brain Atlas including several monkey brain maps (Paxinos et al., 1999; Lewis and Essen, 2000; Kötter and Wanke, 2005; Markov et al., 2011, 2014).

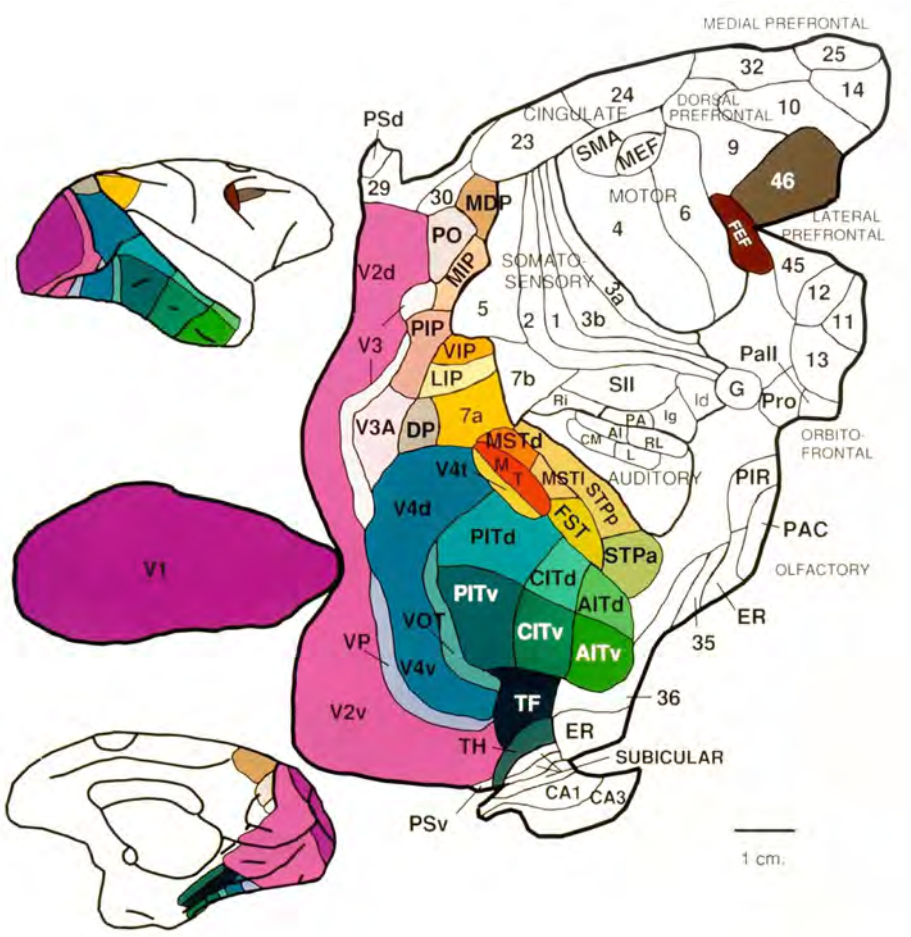


Figure 1-23 The areas in a flat brain charted by Felleman and Van Essen. (Felleman and Van Essen, 1991)

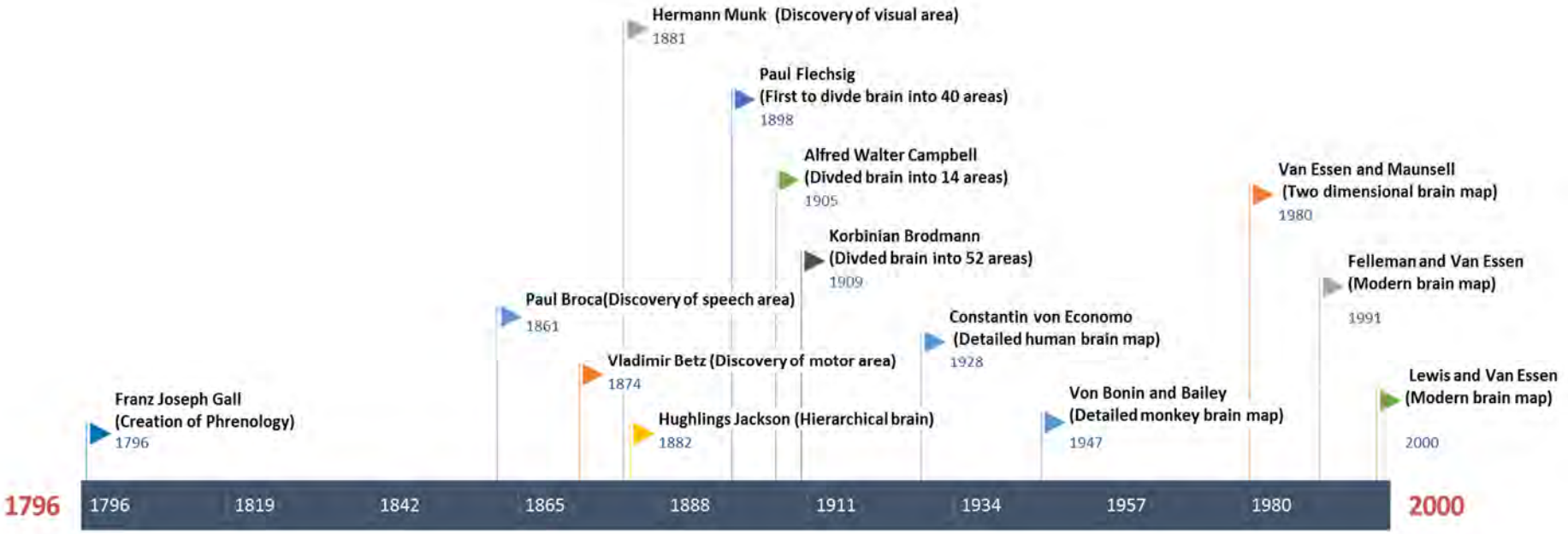


Figure 1-24 The brief history of our understanding of the macro-structure of the neo-cortex

Micro-structure of neocortex

In the parallel time period, based on the cell types and properties in different brain areas, the development of cytoarchitecture helped the idea of functional organization of the brain. The French neuroscientist Jules Baillarger (1809-1890) began the first scientific investigation on the structure of the grey matter of the cortex and divided the cerebral cortex into 6 layers (~1840, this 6-layers structure is not well described due to the crude observation method). Theodor Meynert (1833-1892) first found regional variations (dividing allocortex from neocortex) in different cortical regions (~1867), with a detailed account for the structure in cerebral cortex in general by describing the areas we know today as the visual areas. Vladimir Betz also described different layers of the cortex in details (~1881). His observation of the lamination was translated (Bailey and Bonin, 1951) as:

The cortical substance consists of five different layers which, from without inward, are superimposed on one another in the following manner: The first layer consists of a thick network called neuroglia in which are strewn, here and there, small granular bodies. The second layer contains, besides the neuroglia (which, moreover, all the layers contain) pyramidal cells not too large which, not very near each other, have their apices directed toward the first layer, the base toward the bottom. The third layer is composed of the same pyramidal cells, only two or three times larger, but in compensation less numerous and further apart from one another. The fourth layer, called the granular layer, consists of small, round or elliptical cells. The fifth layer finally consists of specific fusiform cells. This structure of five layers may be considered as the general type

of the cortical substance.

Ramon y Cajal then used the Nissl method to identify in stained material a 9-layers structure (Cajal, 1899) including:

1. *plexiform layer (layer of horizontal cells)*
2. *layer of small pyramids*
3. *layer of medium pyramids*
4. *layer of large stellate cells*
5. *layer of small stellate cells*
6. *layer of small pyramids with arcuate axons*
7. *layer of giant pyramids (solitary cells of Meynert)*
8. *layer of large pyramids with arcuate and ascending axons*
9. *layer of triangular and fusiform cells*

Campbell liked to name the layers or area with description. He claimed 7 layers laminar structure (Campbell, 1905):

1. *Plexiform Layer*
2. *Layer of Small Pyramidal Cells*
3. *Layer of Medium-Sized Pyramidal Cells*
4. *External Layer of Large Pyramidal Cells*
5. *Layer of Stellate Cells*
6. *Internal Layer of Large Pyramidal Cells*
7. *Layer of Spindle-Shaped Cells*

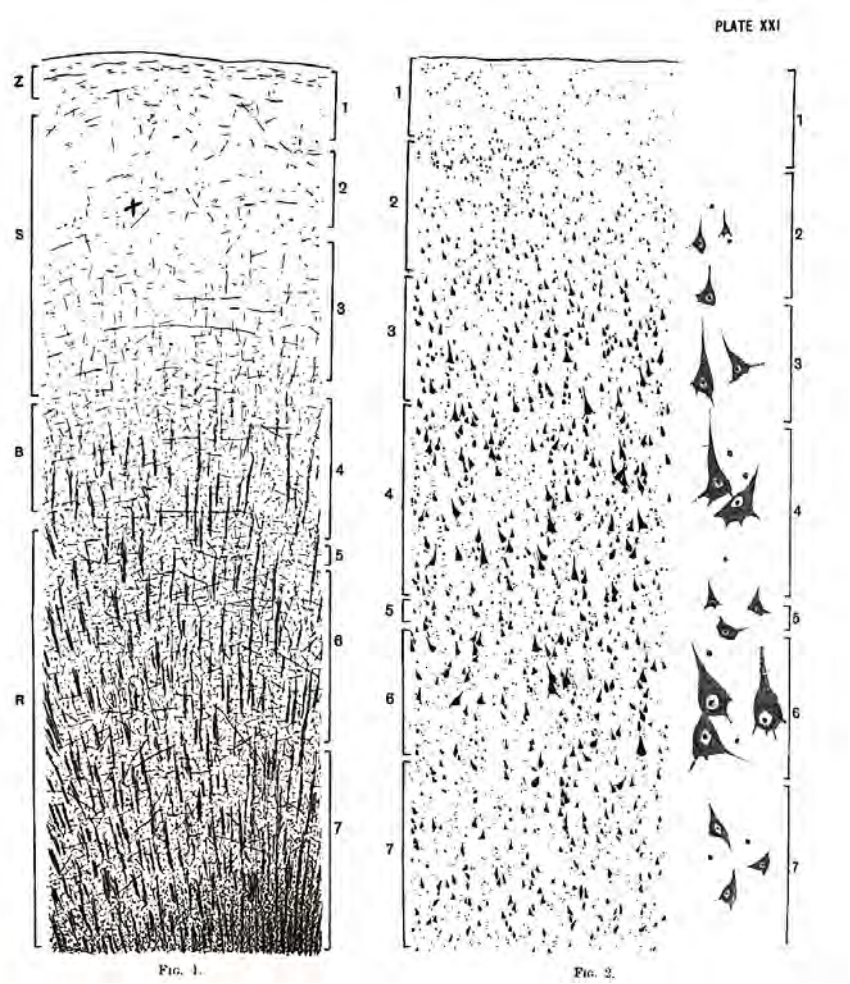


Figure 1-25 The 7 layers of neocortex charted by Campbell. (Campbell, 1905)

Brodmann not only created his famous Brodmann areas, but also created the 6 layers of laminar structure as we know (Brodmann, 1909). He liked to name the areas and layers with numbers. His famous classification in monkey visual cortex (Brodmann, 1905) could be translated (Billings-Gagliardi et al., 1974) as:

I. *Lamina zonalis - the narrow cell-free cortical border.*
II. *Lamina granularis externa - very feebly developed and hardly separable from the adjacent pyramidal layer. Fetal brains show this layer best.*

III. *Lamina pyramidalis - pyramidal cells are located superficially; somewhat larger pyramidal cells are found only in deeper parts.*

IV. (a) Lamina granularis interna superficialis - stands out in the photographs as a distinct dark cell stripe. At higher magnification many little round cells (so-called granules) can be recognized, apart from larger slender star- and pyramid-shaped cells. (b) Lamina (granularis interna) intermedia - contains the stripe of Gennari in fibre preparations. In cell preparations the layer stands out as a wide, cell-poor band containing single large cells, which arrange themselves here and there in the middle into a somewhat denser cell layer. (In other species, namely Cebus capucinus, these large ceils of the lamina intermedia form a distinct, compact cell layer in the middle of IVb, so that here one can again make three subdivisions.) (c) Lamina granularis interna profunda - This is the most cell-rich and, because of this, the darkest, most prominent layer in any cortical cross section. It contains predominantly densely packed granules. With exacting study, particularly with higher magnification or in Bielschowsky preparations, one can also differentiate two layers within this layer – a darker, outer layer composed of granules and large pleomorphic cells, and a light, somewhat thinner, inner layer possessing almost exclusively granules. In other brains, especially of Cebus capucinus, this difference is so significant and conclusive that one can demonstrate two separate layers. However, in the species studied here we ought to retain the layering system set forth for man since this state of things is only hinted at.

V. *Lamina ganglionaris - the most cell-poor and therefore the lightest layer of Area 17. It contains in its deeper portion (bordering on layer VI) a few scattered enormous pyramidal cells, the so-called solitary cells of Meynert.*

VI. *Lamina multiformis - can be more clearly subdivided than in man into two subdivisions: (a) Lamina triangularis, a darker outer layer containing mostly larger cells, and (b) Laminafusiformis, the lighter cell-poor inner layer, or the true spindle-cell layer, which stands out sharply against the white matter.*

Comparing different kinds of classification, we may notice that even if there are differences between different methods of classification, the common structure in these observations are similar, from outside to inside: one cell-free layer, one pyramid cells layer, one stellate cells layer, another pyramid cells layer and one triangular/fusiform cells layer (See more about the lamination classifications of different authors in Appendix). I think we could use a much simpler way to describe the neocortex lamination: the supragranular layer, the granular layer and the infragranular layer.

Same as for the macro-structure of the neocortex, since we still use brodmann's classification as lamination structure of neocortex, we can consider the brodmann's micro-structure map is the state-of-the-arts in this area. The more modern types of lamination classifications usually include the advancement of techniques. For example, Von Bonin used a combination of Nissl, Bodian, and Golgi methods to examine the visual cortex (Bonin, 1942). Garey used light and electron microscopic to study the visual cortex (Garey, 1971). Fatterpekar et al. used MR microscopy (9.4T MRI device) to investigate the human cerebral cortex (Fatterpekar et al., 2002).

Brodmann (1902) (*25)	Meynert (1868)	Betz (1881)	Hammarberg (1895)	Schlapp (1898)	Bolton (1900) & Mott (1907) (*29)	R. y Cajal (1902)	Campbell (1905) (*28)
I. Lamina zonalis	1. Molecular layer	1. Neuroglial layer	1. Molecular layer	1. Tangential fibre layer	1. Outer layer of nerve fibres	1. Plexiform layer	1. Plexiform layer
II. L. granularis externa	2. Small pyramids	2. Small pyramids	2. Small pyramids (II & III)	2. Outer polymorphic cells	2. Small pyramids	2. Small pyramids (II & III)	2. Small pyramids (II & III)
III. L. pyramidalis				3. Pyramidal cells			
IVa. L. granularis interna superficialis	3. Outer granular layer	3. First granular layer	3. Small cells (IVa) and solitary cells (IVb)	4. Granular layer	3a. Outer layer of granules	3. Medium pyramids (IIIb and IVa)	3. Medium pyramids (IIIb and IVa)
IVb. L. granularis interna intermedia (stria of Gennari)	4. Outer intermediate granular layer and solitary cells	4. Longitudinal fibre layer	4. Dense single cells (IVb)	5. Solitary cells	3b. Middle layer of nerve fibres (Gennari)	4. Large stellate cells (IVb)	4. Stellate cells (IVb)
IVc. L. granularis interna profunda	5. Middle granular layer	5. Second granular layer	5. Small granules (IVc)	6. Granular layer	3c. Inner layer of granules	5. Small stellate cells (IVc)	5. Small stellate cells (IVc)
V. L. ganglionaris	6. Inner intermediate granular layer and large solitary cells	6. Second fibre layer	6. Solitary cells (V)	7. Cell-poor layer	4. Inner layer of nerve fibres	6. Small cells with arciform axons (IVc)	6. Solitary cells
		7. Large pyramids				7. Giant pyramids	
VI. L. multiformis	7. Inner granular layer		7. Polymorphic cells			8. Large cells with ascending axons	
VIa. L triangularis		8. Spindle cells		8. Inner polymorphic layer	5. Polymorphic cells		7. Spindle cells
VIb. L. fusiformis	8. Spindle cells		8. Spindle cells			9. Spindle cells	

Table 1-4 Laminations of the visual cortex according to different authors (Brodmann, 1909)

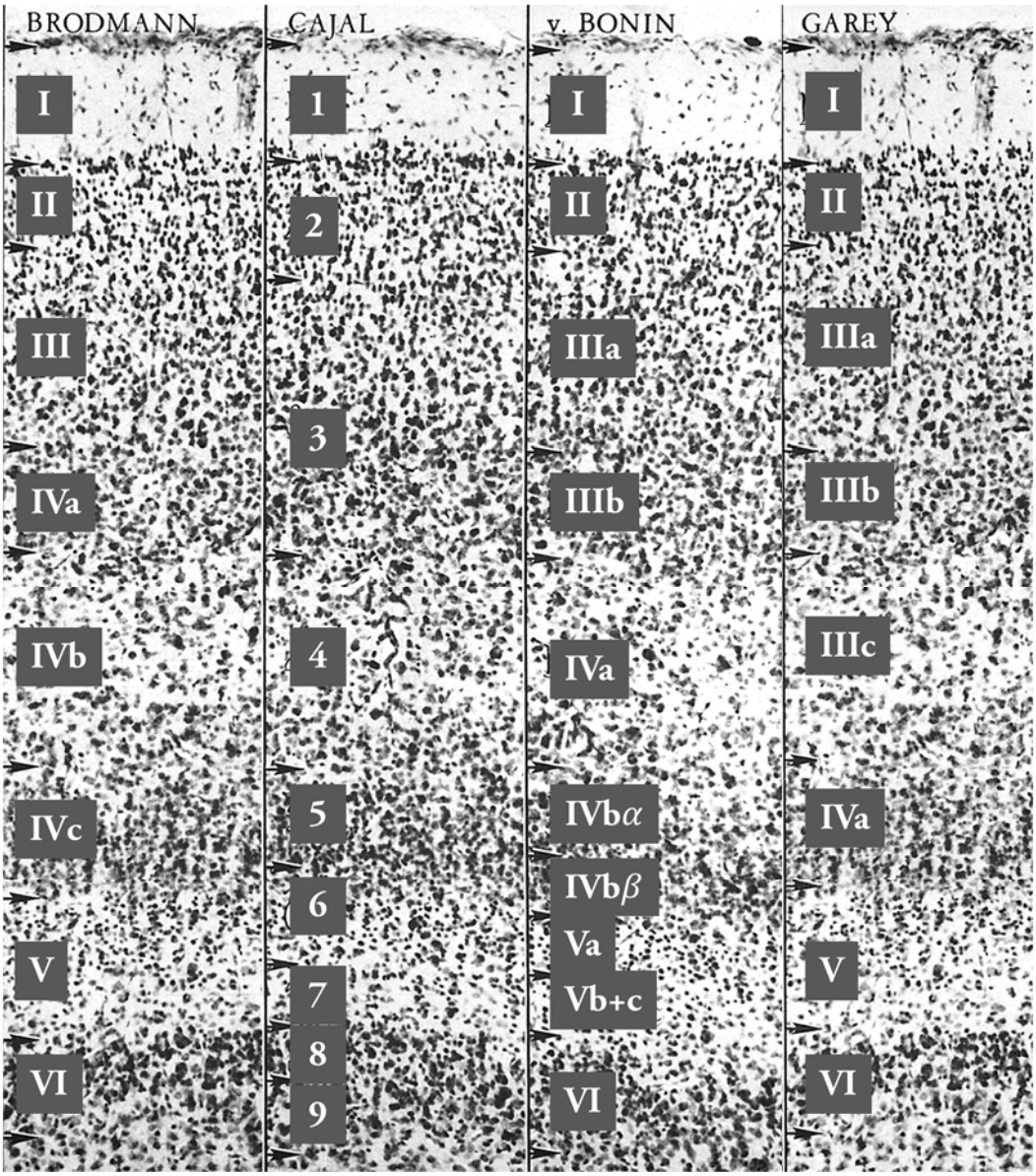


Figure 1-26 Laminations of the neocortex according to different authors. (Billings-Gagliardi et al., 1974)

On the lamination classification of neocortex, I think we should notice the following two points:

1. The neocortex is a biological tissue and the laminar classification

is more about the degree of concentration of certain types of neurons (or shapes of neurons), thus, there are no hard lines between different layers. The variances are also big between different cortical areas. In vivo recording, the depths of the electrodes were recorded, but the layers of recordings were reported based on experience. Thus, as Douglas and Martin pointed out, “the six layered neocortex is something of a unicorn” (Douglas and Martin, 2007).

2. The neurons in one layer can receive input from another layer. Since the thickness of human neocortex is from 1 to 4.5 mm (Order of magnitude: 10^0 mm) (Fischl and Dale, 2000), while the typical length of primary apical dendrite of pyramidal neuron have an order of magnitude of 10^{-1} mm, there is a very large possibility that the dendrite could cross more than one layer. Thus, we should be careful with the so called “laminar computation” (see (Larkum, 2013)) and the observations of different input strength to different layers since the deeper layer could potentially receive the input from the shallower layers.

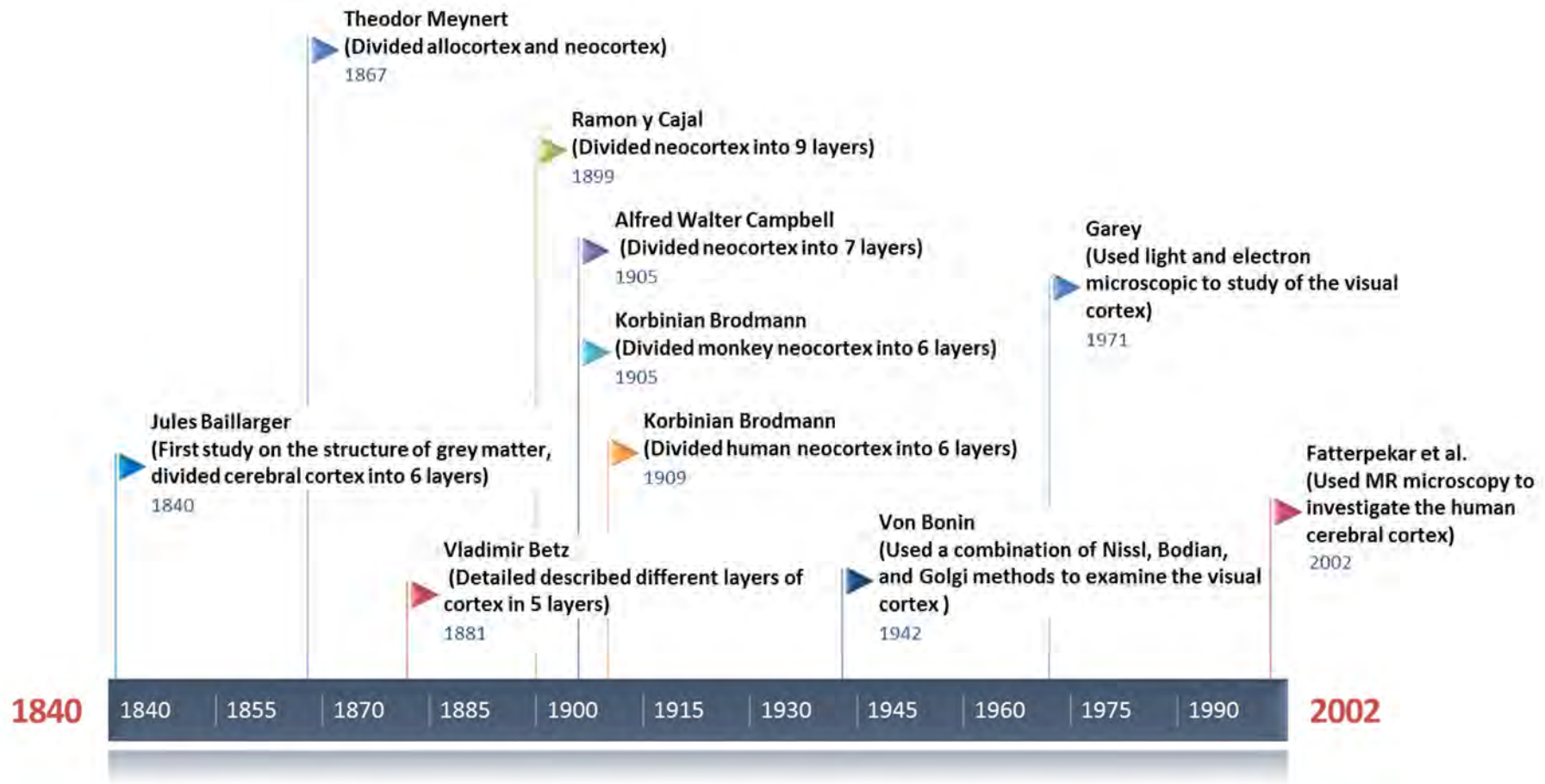


Figure 1-27 The brief history of our understanding of the micro structure of the neo-cortex

Connectivity of neocortex

The brain is heavily connected. In humans, the volume of white matter reaches about 80% of the total volume (Zhang and Sejnowski, 2000), which contains mostly the glial cells which produce the myelin to speed up the connection between different areas. There is a strong correlation (see appendix) between the volume of gray matter and white matter, which suggests this heavy connection between different areas is not only applicable to human, but universal for different species (Zhang and Sejnowski, 2000).

The connectivity of the neocortex was also heavily studied with different methods and is still on going. The most well studied connectivity is between different visual areas and it has been widely accepted that the brain employs a hierarchical structure to implement its different functions, thus, the connectivity between different areas usually characterized as "feedforward" or "feedback". Furthermore, the studies on these different directions suggest different connection patterns.

Hierarchical brain

Hughlings Jackson first proposed the idea of a hierarchical brain based on the evolution theory in 1882 (York and Steinberg, 2011). He stated:

The higher the centre the more numerous, different, and more complex, and more special movements it represents, and the wider region it represents-evolution. The highest centres represent innumerable, most complex and most special movements of the organism, and…each unit of them represents the organism differently. In consequence, the higher the centre the more numerous, complex and special movements

of a wider region are lost from a negative lesion of equal volume-dissolution. (Jackson, 1882)

This proposal is more like a general claim rather than a scientific conclusion since little evidence was given. Nowadays, the hierarchical brain evidence mostly comes from the visual system: Hubel and Wiesel showed a progressive increase in the complexity of the cat visual cortex (Hubel and Wiesel, 1962). Other studies showed that the connections from area 17 mostly rise from the supragranular layers and terminate in the layer 4 of the target areas and connections raised from the infragranular layers usually have terminals that avoid layer 4, see more from the review by Salin and Bullier (Salin and Bullier, 1995). Starting from these results, researchers began to assign the different brain areas with different brain hierarchy levels (Rockland and Pandya, 1979; Friedman, 1983; Van Essen and Maunsell, 1983). This notion was confirmed by the famous paper by Felleman and Van Essen (Felleman and Van Essen, 1991) which is published in the very first issue in the journal Cerebral Cortex .

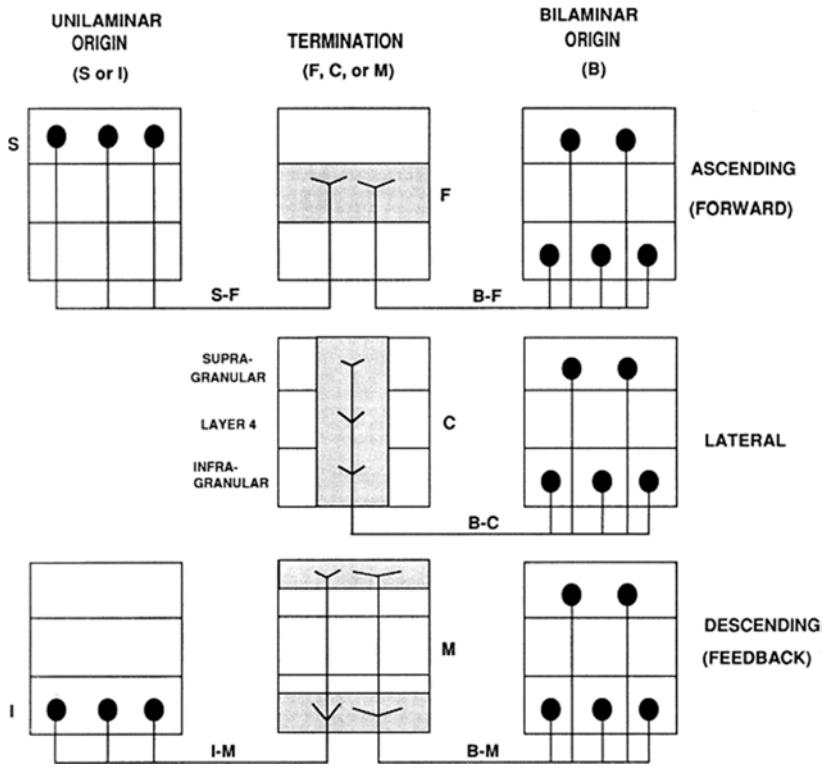


Figure 1-28 The definition of feedforward and feedback connections. The nearby areas usually have a bilaminar origin and the distant areas usually have a unilaminar origin

In their paper, they first defined 32 visual areas using their two dimensional map and then showed the connectivity between different areas. Felleman and Van Essen reported a 32 × 32 connectivity matrix with 305 known projections out of 992 possible pathways. Then they took advantage of the rising and terminal layer of the connection (feedforward connections rising from supragranular layers or supragranular & infragranular layers and terminate in layer 4, feedback connections rising from infragranular layers or supragranular & infragranular layers and terminate in all layers except layer 4) and created “feedforward” and “feedback” connections (Felleman and

Van Essen, 1991). The connections in neighboring areas (e.g. V1 and V2) usually have a bilaminar origin, while the connections in distant areas (e.g. V1 and MT) usually have a unilaminar origin (Salin and Bullier, 1995).

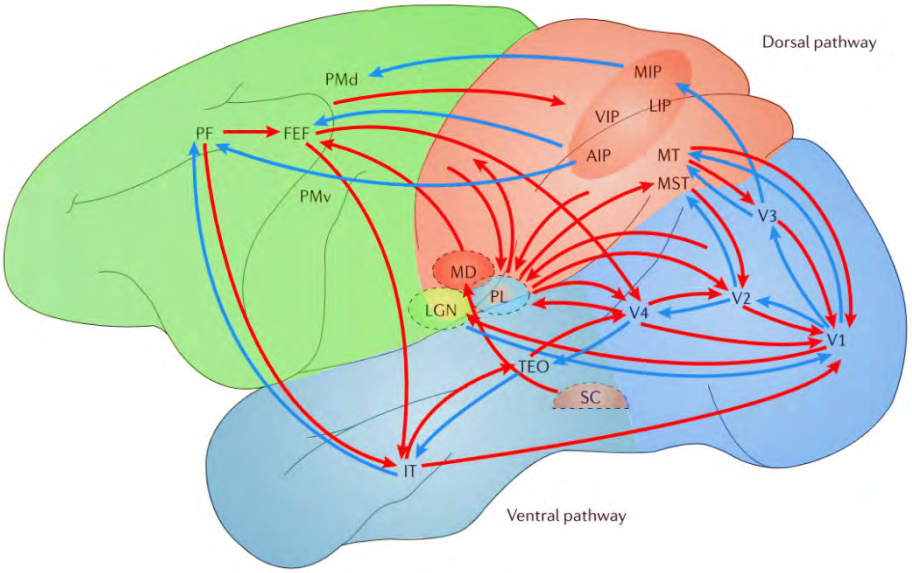


Figure 1-29 A simplified connectivity map with the feedforward and feedback connections (Blue arrows: feedforward connections, red arrows: feedback connections).

Subsequent studies were done using improved tracers and began to analyze the weight between different areas. Recent review showed that the improved tracer helped to find 36% new connections between different areas (Markov et al., 2013a). Furthermore, based on the idea of neighboring areas having bilaminar origin, while distant area has a unilaminar origin, researchers found a distance rule. Based on this distance rule, researchers checked the fraction of supragranular labelled neurons and modified the hierarchy order of different brain areas (Barone et al., 2000; Markov et al., 2011, 2013b).

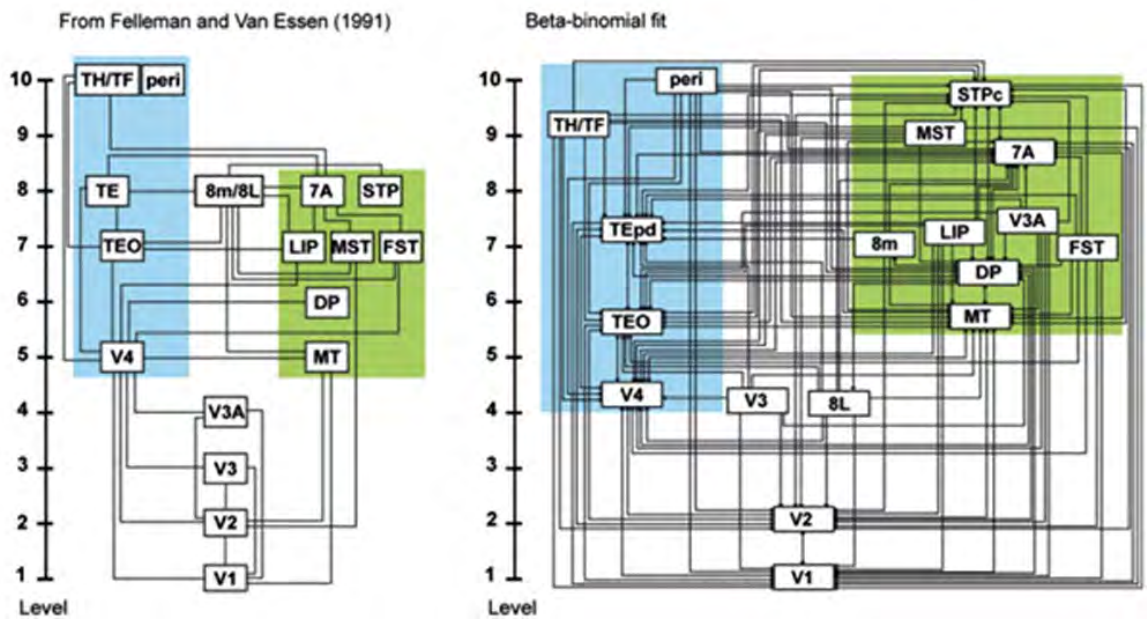


Figure 1-30 Comparison of Felleman and Van Essen's hierarchy order (left) and hierarchy order based on the unilaminar and bilaminar distant rule (right). From (Barone et al., 2000).

Roles of feedforward and feedback connections

Since we know the existence of two types of neuronal functions: excitatory and inhibitory, and brain areas are heavily connected with feedforward and feedback connections, it is interesting and important to know the roles of feedforward and feedback connections: are they excitatory or inhibitory? To answer this question, we need to consider the types of the projection neurons, the target neurons and the experimentally observed feedforward and feedback effects.

Projection neurons:

From the beginning of this thesis (page 14-18), we know that different types of neurons have different shapes and the pyramidal neurons are

the only type of neuron containing a long axon. Since the distance between different areas are far (from 6 mm to 45.6 mm in monkey, see details in Markov et al., 2013b, with the order of magnitude of 10^1 mm), it is physically impossible for other types of neurons to play the role of projection neuron. Thus, the projection neurons for both feedforward and feedback connections are excitatory neurons. Experimental evidence supported the same conclusion as this physical limitation suggested (Johnson and Burkhalter, 1996).

Target neurons:

The target neurons are not much more diverse than the projection neurons. From 1970s, scientists begin to use electron microscope (EM) and tracer to investigate the connections. The electron microscope uses a beam of accelerated electrons as a source of illumination and can achieve the resolution to 50 pm (magnifications of up to 10,000,000x). Thus, it is possible to directly observe the synaptic connections. For the tracer, two methods are usually used: lesion and HRP. The lesion of one particular area (such as LGN) can lead to a degeneration (which is visible using the EM) of their synapses (such as the synapses connected to V1). The other way is using a special kind of method (iontophoretic delivery) to deliver the HRP or other kinds of tracer to neurons which can lead to a Golgi-stained effect (which is also visible using the EM). Under the EM, by simply counting the numbers of synapses in different target positions, we can know the properties of the synapses.

Three possible synapse targets can be observed using the EM: dendritic spines, dendritic shafts and cell bodies (soma). Since only the excitatory neurons have dendritic spines, we can know the amount of neurons that are excitatory. In some studies, we can know a detailed ratio of the excitatory and inhibitory neurons: for the synapses that are connected

to the shafts (also soma, but there is few), we can test whether the post-synaptic neuron is GABA positive or not. By multiplying this ratio and the percentage of the shaft/soma connections of the synapses, we can know the percentage of inhibitory targets.

We show the data from different studies in the following table. For the thalamic feedforward (LGN to cortex), studies show a mostly excitatory targeting profile: About 80% of the connections from LGN to V1 target on the dendritic spines. About 20% of the connections target on the dendritic shafts. Single digit percentage of the connections target on the soma. Further examinations on the positivity of the GABA show that about 4-9% of all the connections actually target on inhibitory neurons.

For cortical feedforward connections (e.g. from V1 to LM in mouse), similar to the thalamic feedforward connections, about 80% of the connections target on the dendritic spines with about 20% on the dendritic shafts. But the confirmed GABA positive targeting ratio is higher: 10-15% of all the connections actually target on the inhibitory neurons.

For cortical feedback connections (e.g. from LM to V1 in mouse), about 98% of the connections targets on the dendritic spines with about 2% of the connections targeting on the dendritic shafts. The confirmed GABA positive targeting ratio is also very low: ~2%. This suggests an excitatory profile for feedback connections.

Additionally, for callosal connections (one hemisphere to another), studies showed that more than 96% of the connections are targeting on dendritic spines, regardless of the areas (the same or different areas) they are targeting on. For intrinsic connections, 90% of synapses that connect to the superficial layers' neurons in rat target on the dendritic spines. In cat motor cortex, 60% layer 4 neurons' synapses that connect

to the layer 2/3 target on the dendritic spines. In cat V1, 28% layer 6 neurons' synapses that connected to the layer 4 target on dendritic spines.

Furthermore, in mouse, Shao and Burkhalter found that thalamocortical inputs, feedforward and local connections inputs within V1 evoked monosynaptic excitatory postsynaptic potentials (EPSPs), followed by disynaptic, hyperpolarizing inhibitory postsynaptic potentials (IPSPs) (10/11, 91% for thalamocortical input; 17/19, 89% for feedforward; 12/13, 92% for local connections). However, for the feedback connections, only 13/58, 22% connections showed such a profile (45/58, 78% of the cells in V1 activated by feedback input showed monosynaptic responses that were depolarizing only). This result suggests that there is a stronger thalamocortical, feedforward, and local connection input into the inhibitory neurons than the feedback input.

In summary, the profile for different kinds of connections are: a strong excitatory feedforward connection, an even stronger excitatory feedback connection, and complicated intrinsic connections.

Table 1-5 The targeted positions for different connections

Study	Species	From area	Observation site	Method	Connection number	Targeting on			Connected to GABA positive neurons
						Spine	Shafts	Soma	
(Peters and Feldman, 1977)	Sprague-Dawley albino rats	LGN	V1 Layer 4	LGN Lesions and observe the degeneration in V1	256	213 (83%)	37 (15%)	6 (2%)	
		All	V1 Layer 4	Direct observation	500	Asymmetric, 372 (74%). Symmetric, 5 (1%)	Asymmetric, 62 (12%). Symmetric, 57 (12%)	Symmetric, 4 (1%)	
(Garey and Powell, 1971)	Monkey	LGN	Area 17 Layer 4	LGN of Lesion	294	247 (84%)	40 (14%)	7 (2%)	
	Cat	LGN	Area 17 Layer 4		330	273 (83%)	48 (14%)	9 (3%)	
	Cat	LGN	Area 18 Layer 4		255	179 (70%)	50 (20%)	26 (10%)	
	Cat	LGN	Area 19 Layer 4		148	126 (85%)	21 (14%)	1 (<1%)	
(Freund et al., 1985)	Cat	LGN	Area 17 Layer 4, occasional layer 3	Iontophoresis horseradish peroxidase (HRP)	306	X (~90%),	X (~5%),	X (~5%),	
						Y (~80%)	Y (~15%)	Y (~5%)	
		LGN	Area 18 Layer 4			Y (~70%)	Y (~25%)	Y (~5%)	
(Freund et al., 1989)	Monkey	LGN	Area 17	HRP		PA (68.9%)	PA (33%)	PA(3.1%)	PA (2/53), 3.8%.
						MA (51.5%)	MA (47.2%)	MA(1.3%)	MA (14/229), 6.2%
		LGN	Area 17			PA (55.4%)	PA (43.1%)	PA (1.5%)	PA (4/65), 4.2%

Study	Species	From area	Observation site	Method	Connection number	Targeting on			Connected to GABA positive neurons
						Spine	Shafts	Soma	
						MA (64.3%)	MA (35.7%)	MA (0%)	MA (12/126),9.5%
(Lowenstein and Somogyi, 1991)	Cat	V1	PMLS (middle layers)	Iontophoretically delivered phosphate-buffered saline	190	158 (83%)	32 (17%)		14.45%, based on the percentage of the test GABA+ shafts (11/13, 85%)
		All	PMLS (middle layers)		893	634 (71%)	258 (29%)		7.83%, based on the percentage of the test GABA+ shafts (71/258, 27%)
(Kisvarday et al., 1986)	Cat	Area 17 Layer 3	Area 17 layer 3 same column	HRP	191	62 (86.1%)	9 (12.5%)	1 (1.4%)	4.55%, based on the percentage of the test GABA+ shafts (4/12, 33.3%)
		Area 17 Layer 3	Area 17 Layer 5 same column			31 (86.1%)	5 (13.9%)		
		Area 17 Layer 3	Area 17 layer 3 different column			61 (87.1%)	9 (12.9%)		
		Area 17 Layer 3	Area 17 Layer 5 different column			11 (84.6%)	2 (15.4%)		
(Bueno-Lopez et al., 1989)	Cat	All	Area 17 Layer 4		794	421 (53%)	365 (46%)		9.44%, based on the percentage of the test GABA+ shafts (75/365, 20.5%)
(White and Czeiger, 1991)	Mouse	Callsoal axon	Intrinsic terminals	HRP	1215	1174 (96.6%)	41 (3.4%)		
		Callsoal axon	Extrinsic terminals		277	269 (97.1%)	8 (2.9%)		
		All	Area 1 layer 2 and 3		398	331 (83.2%)	67 (16.8%)		
		All	Area 40 layer 2 and 3		388	309 (79.6%)	79 (20.4%)		

Study	Species	From area	Observation site	Method	Connection number	Targeting on			Connected to GABA positive neurons
						Spine	Shafts	Soma	
(McGuire et al., 1991)	Monkey	Area 17 layer 3B	Area 17	HRP	117	89 (76%)	28 (24%)		
		All excitatory (asymmetric)	Area 17 layer 3		267	181 (68%)	83 (31%)	3 (1%)	33/300, 11%, based on the asymmetric and symmetric synapse number
		All inhibitory	Area 17 Layer 3		33	11 (33%)	17 (52%)	5 (15%)	
		(Symmetric)							
(Johnson and Burkhalter, 1996)	Long Evans rats	All	Area LM Layer 1	Anterograde axonal tracing with the kidney bean lectin haseolus vulgarisleucoagglutini		44 (89.8%)	5 (10.2%)		10.2%, based on the percentage of the test GABA+ shafts (5/5, 100%)
		All	Area LM Layer 2/3			67 (88.2%)	9 (11.8%)		7.99%, based on the percentage of the test GABA+ shafts (6/9, 67.7%)
		All	Area LM Layer 4			14 (87.5%)	2 (12.5%)		6.25%, based on the percentage of the test GABA+ shafts (1/2, 50%)
		All	Area 17 Layer 1			47 (87%)	7 (13%)		7.42%, based on the percentage of the test GABA+ shafts (4/7, 57.1%)
		All	Area 17 Layer 2/3			41 (85.4%)	7 (14.6%)		12.5%, based on the percentage of the test GABA+ shafts (6/7, 85.7%)
		Area 17	Area LM layer 1			13 (100%)			

Study	Species	From area	Observation site	Method	Connection number	Targeting on			Connected to GABA positive neurons
						Spine	Shafts	Soma	
		Area 17	Area LM layer 2/3			80 (89.9%)	9 (10.1%)		10.1%, based on the percentage of the test GABA+ shafts (9/9, 100%)
		Area 17	Area LM Layer 4			26 (86.7%)	4 (13.3%)		13.3%, based on the percentage of the test GABA+ shafts (4/4, 100%)
		Area LM	Area 17 Layer 1			72 (100%)	0 (0%)		
		Area LM	Area 17 Layer 2/3			110 (97.3%)	3 (2.7%)		2.7%, based on the percentage of the test GABA+ shafts (2/2, 100%)
(Johnson and Burkhalter, 1997)	Long Evans rats	Area 17 feedforward connections	Area 17 Layer 1 collateral connections	biotinylated dextran amine labelling		43 (95.6%)	2 (4.4%)		
		Area 17 feedforward connections	Area 17 Layer 2/3 collateral connections			16 (88.9%)	2(11.1%)		
		Area LM feedback connections	Area LM Layer 1 collateral connections			29 (93.5%)	2 (6.5%)		
		Area LM feedback connections	Area LM Layer 2/3 collateral connections			5 (100%)	0		
(Keller and Asanuma, 1993)	Cats	Motor cortex Layer 4	Motor cortex Layer 2/3	Neurobiotin	161	101 (63%)	52 (32%)	8 (5%)	
		Motor cortex Layer 4	Motor cortex Layer 2/3 pyramidal neuron		20	18 (90%)	2 (10%)		

Study	Species	From area	Observation site	Method	Connection number	Targeting on			Connected to GABA positive neurons
						Spine	Shafts	Soma	
		Motor cortex Layer 4	Motor cortex Layer 2/3 non-pyramidal neuron		30	2 (6.7%)	20 (66.7%)	8 (22.6%)	
(McGuire et al., 1984)	Cats	Area 17 Layer 6	Area 17 Layer 4	HRP	151	43 (28%)	108 (72%)		
(Gabbott et al., 1987)	Cat	Area 17 Layer 5	Area 17 layer 4	HRP	49	~96%	~4%		
		Area 17 Layer 5	Area 18 layer 4		6	100%			
		Area 17 Layer 5	Area 17 layer 5		77	~80%	~20%		
		Area 17 Layer 5	Area 18 layer 5		20	~80%	~20%		
		Area 17 Layer 5	Area 17 layer 6		75	~65%	~35%		
		Area 17 Layer 5	Area 18 layer 6		39	~70%	~30%		

Observed effects:

There are many methods to observe the effects of feedforward and feedback connections (e.g. lesion, cooling down or optogenetic activation or deactivation in higher or lower areas). The underlying principle is simple: modify the activity of higher or lower area and observe the target areas' response.

For the feedforward connections, monkey recording studies by Girard et al showed that the neurons in V2 (~100%), V3a (~70%), V3 (~100%) and V4 (~100%) were not responding to the visual stimuli after reversibly inactivating V1 by cooling(Girard and Bullier, 1989; Girard et al., 1991a, 1991b). These effects only worked on the neurons with receptive fields which were included in the visual field region coded by the inactivated zone, the neurons outside this region remained visually responsive. However, for V5/MT, Girard et al showed that most of the neurons (~80%, ~20% not responding) were still responding to the visual stimuli when cooling V1 (Girard et al., 1992). Lesion studies on V1 also confirmed this observation: 66% of neurons in macaque MT still respond to visual stimulation 5-6 weeks after a lesion of V1 (Rodman et al., 1989).

For the feedback connections, the observed effect is similar to the feedforward connections. In anesthetized animals, Sandell and Schiller showed that cooling down V2 would lead to an activity drop in V1 (~86%) (Sandell and Schiller, 1982). Hupé et al. showed that in most cases, cooling down MT would lead to an activity drop in V1, V2 and V3 (~84%), while in one extreme case of very low saliency stimuli, V3's activity increased during the cooling(Hupé et al., 1998). Hupé et al. also showed similar results in V1 as Sandell and Schiller: tuning down V2 activity with GABA resulted in an activity drop in V1 (100%, but with only 6 neurons) (Hupe et al., 2001). Wang et al. showed that cooling PTV in cat could

reduce the activity in V1 (~81%) (Wang et al., 2010).

However, recent studies on awake animals showed that, when cooling down V2/V3, about half of affected neurons inside the classical receptive field in V1 increased their activity (~52% increased activity, ~48% decreased activity), and when including the surrounding neurons of the classical receptive field, most neurons increased their activity (~89% increased activity, ~11% decreased activity) (Nassi et al., 2013). In mouse, Zhang et al. showed that focal activation of Cg (frontal cortical area) axons in V1 caused a response increase at the activation site but a decrease at nearby locations (center-surround modulation) (Zhang et al., 2014).

In summary, both feedforward and feedback have mostly an excitatory effect. But in certain conditions, they can also have an inhibitory effect. The excitatory and inhibitory effects could have different reasons, and affect different regions.

Table 1-6 Effect of inactivation of the feedback from higher neuronal area. Modified from (Nassi et al., 2013)

						CRF-only stimulation			CRF + surround stimulation		
Study	Species	Anesthetic	Inactivation method	Pathway	Stimulus	% affected	% less activity[a]	% more activity[a]	% affected	% less activity[a]	% more activity[a]
Nassi et al.(2013)	Macaque	None	Cooling	V2/V3 → V1	Sinusoidal grating	32% (21 of 66)	48% (10 of 21)	52% (11 of 21)	67% (44 of 66)	11% (5 of 44)	89% (39 of 44)
Sandell and Schiller (1982)	Saimiri	N_2O + halothane thiopental	Cooling	V2 → V1	Single bar	32%[b] (21 of 66)	86%[b] (18 of 21)	14%[b] (3 of 21)	Not tested		
Hupé et al. (2001a)	Macaque	N_2O + sufentanil	GABA	V2 → V1	Array of bars	10% (6 of 61)	100% (6 of 6)	0% (0 of 6)	No effect on surround modulation Specific breakdown not reported		
Hupé et al. (1998)	Macaque	N_2O + sufentanil	Cooling	MT → V1/V2/V3	Array of bars	40% (61 of 154)	84% (51 of 61)	16% (10 of 61)	Population response increase[c] Specific breakdown not reported		
Wang et al. (2010)	Cat	N_2O + halothane	Cooling	PTV → V1	Sinusoidal grating	39% (16 of 41)	81% (13 of 16)	19% (3 of 16)	76%[b] (31 of 41)	81%[b] (25 of 31)	19%[b] (6 of 31)

[a]"Less" and "more" indicate changes in V1 response during inactivation of feedback.
[b]Values were not tested for statistical significance.
[c]Observed for V3 and low-contrast stimuli only.

The Convergence and Divergence

One other aspect of the connectivity between different cortical areas is by measuring the convergence and divergence. Since most of the cortical areas connect to each other in a reciprocal fashion (one area both sends and receives signals from the other area), it would help us a great deal to know if there are differences between the feedforward and feedback connections. The studies about the convergence and divergence have demonstrated a rather clear image: feedforward connections are more convergent and feedback connections are more divergent.

The most direct evidence for this conclusion may come from studies using retrograde and anterograde tracers between areas 17 and 18 in cat. To determine the feedforward convergence, Ferrer and colleagues used two different retrograde tracers (one with yellow color, and the other with blue color, the tracers go from the target neuron's cell body to the target neuron's dendrite, then to the projection neuron's axon and then to the projection neuron's cell body). They found out that when the boundaries of the dense central cores of two injection sites in area 18 were separated by at least 1.6 mm, the two corresponding distributions of labelled neurons in area 17 were just non-overlapping (Ferrer et al., 1988). This result suggested that the feedforward connections from one point in area 17 should only affect the neurons population with a size 0.8 mm larger in all directions in area 18. However, for the feedback connections in cat, Henry et al. showed that, by using the anterograde tracer (which goes from the target neuron's cell body to the target neuron's axon, then to the projection neuron's cell body), small injections (usually 0.2 – 0.5 mm) in area 18 would lead to a

divergence labelling effect in area 17 (from 3.5 – 6 mm in the mediolateral direction and 7 – 8 mm in the rostrocaudal direction). This result suggests the feedback corticocortical connections are organized in a strongly divergent fashion (Henry et al., 1991).

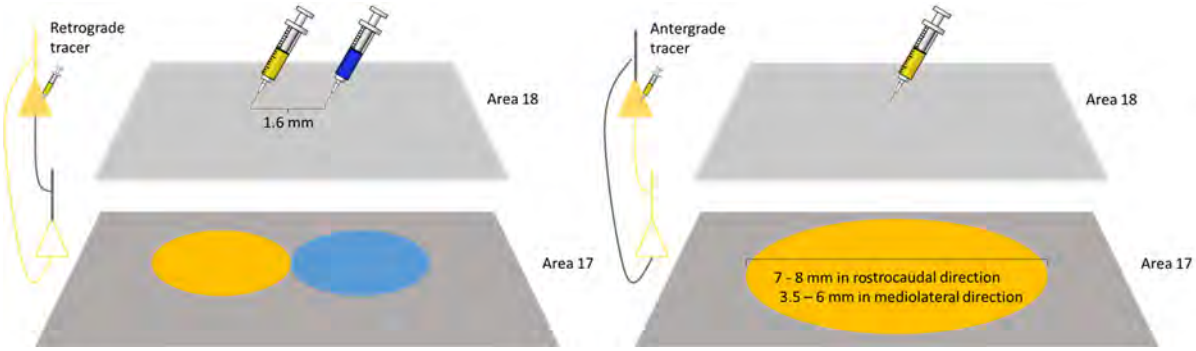


Figure 1-31 The convergence and divergence of feedforward and feedback connections. The left one measured the feedforward divergence and had a maximum of 1.6 mm spread for feedforward connection. The right one measured the feedback divergence and had a result of 3.5 - 8 mm spread for feedback connection.

Another set of evidence for the feedforward and feedback convergence and divergence is axonal bifurcations. It is widely accepted that one important difference between feedforward and feedback is the amount of axonal bifurcations: in cat, there are very little axonal bifurcations (<3%) for the feedforward projections from area 17 to area 18 and 19. But the feedback projections to area 17 contain much more axonal bifurcations (20%-30%) (Salin and Bullier, 1995). Furthermore, it is worth to notice that the proportions of neurons with bifurcations tend to be higher in infragranular than in supragranular layers (Kennedy and Bullier, 1985).

Temporal dynamic of neocortex

Time delay between areas

Time is required for the communication between different areas. It could be simply interpreted as the speed of signal transportation from different areas. However, there are various kinds of ways to measure this time delay and the concept of latency is very easy to be misunderstood. For the information to, travel from one area to another, spike (or electronic signal travelling using the voltage change of the axons) is the only tool. On the other hand, when we present stimuli to subject, at each of the visual areas, there is a significant increase of neurons' firing rate or an event related response (ERP). Here, to better understand this time delay, we divided studies about the time delay into two categories: the axonal conduction delay and the response delay. They represent two fundamentally different measuring methods and functional meanings.

Axonal conduction delay

The axon of the neuron is the carrier of the neuronal signals. By changing the membrane potential, the electronic signal travels through the axon to the synaptic cleft, and to the dendrite of the next neuron. Even though the electronic signal transfer into the chemical signal, the time required for the signal to cross the synaptic cleft is very short (modelling study showed that 50% of the neurotransmitter finished their transmitting and cleared only in 0.05 ms and 90% in 0.5 ms, this time is usually included in the antidromic delay, see more in (Clements et al., 1992; Clements, 1996)). Thus, the time needed to transfer signal from one neuron to another is mainly depended on the time spent on the axon.

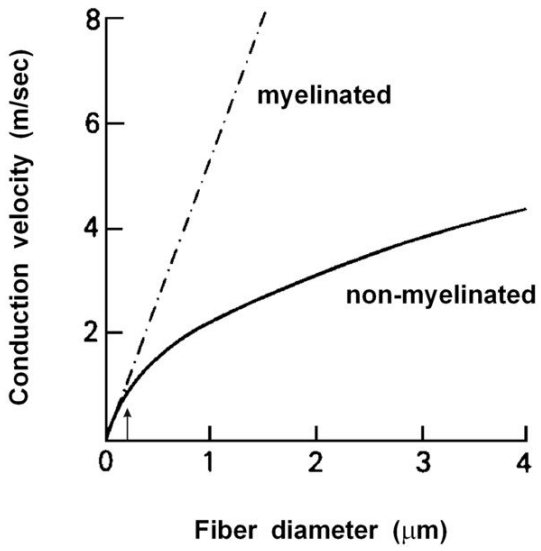


Figure 1-32 Relationship between the conduction velocity and fiber diameter for myelinated and non-myelinated connections.

The axons are like the cables of our brain, and more than 80% of the brain volume is to contain the axons. Two kinds of axons could be found in the brain: non-myelinated and myelinated. The fatty substance of myelin could help to speed up the transportation speed and in mammals, most of the inter-areal connections are myelinated axons (the reason for the white color of the white matter is the myelin). Another factor that affects the axonal conduction speed is the diameter of the axon. The axon diameter and the conduction velocity have a monotonic positive relationship: the thicker the axon, the higher the velocity is.

One way to study this conduction delay is by using the antidromic method which can provide much clear results since the orthodromic recording can not be accurate because of the spontaneous spikes. The distribution of the axonal conduction delays usually has a peak at 1 – 2ms with a long tail until tens of milliseconds (thus the mean value of such distribution can not reflect the features of the data). Connections

between different areas and within different species have different conduction delays, these conduction delays could change from 0.5 ms to more than 30 ms. Some argued that only the short conduction delay can reflect the real conduction delay and define the antidromic delay as “short latency” because the long antidromic delay may be caused by the recording method and noise. For example, since the antidromic delay is measured using the time from the electrical shock from one area to the foot of the antidromic spike. It is possible that the measured spike is not caused by direct electrical shock, but rather the antidromic spike from local neurons and the measured neuron is not directly connected to the electrically shocked area.

Another property of the conduction delay worth noticing is that the conduction delay in one axon is very stable. In other words, the jitter of the conduction delay is very small: the usual criteria for antidromic spike is a latency jitter less than 0.1 ms, for orthodromic spikes, the jitter is 0.3 – 0.5 ms (Girard et al., 2001).

The measured results showed a very fast connection for different areas in cat and monkeys and suggested a similar conduction delay for feedforward and feedback connections. Girard et al showed that the mean delay for feedforward and feedback from V1 to V2 is 1.1 ms and 1.25 ms (Girard et al., 2001). Similar values were showed between V1 and MT (1.3 ms) (Movshon et al., 1996), LIP and FEF (2.3 ms) (Ferraina et al., 2002). However, in rat and rabbit, bigger values were measured: Nowak et al showed that the axonal conduction delay between V1 and V2 is about 5 ms to 6 ms (similar values were obtained for both feedforward and feedback)(Nowak et al., 1997). Swadlow et al showed similar values for V1 and V2 in rabbit (Swadlow and Weyand, 1981).

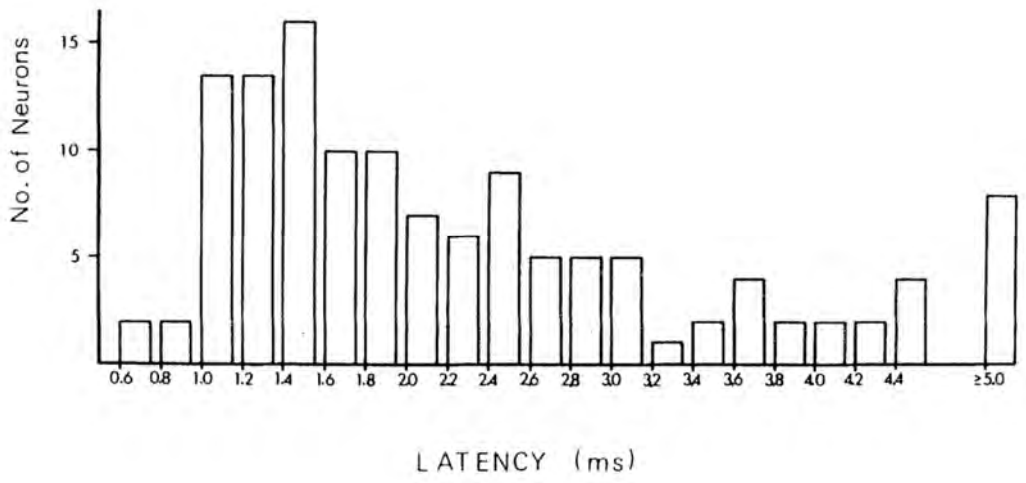


Figure 1-33 Typical distribution of the axonal conduction delay. This is the axonal conduction delay between the S1 and motor cortex, from (Waters et al., 1982)

To sum up, the information transportation between areas is amazingly fast and stable: axonal conduction delays between areas are only 1-2 ms between different cortical areas with the jitter less than 0.1 ms. These numbers are so small and they are even comparable with the industrialized modern information transportation tools: the time delay inside physically connected computer local area network is about 1 ms and the local WIFI time delay is about 10 ms to several hundred ms. These time delays are with a much bigger jitter. There must be a significant functional meaning for a biological organism to spend 80% of its brain volume and very costly materials (myelin) to achieve this industrial level conduction velocity. The advantage should include more than the fast reflex since simple reflex need as long as several hundred milliseconds while the conduction delay is much faster and could be achieved in low level species.

Table 1-7 Axonal conductional delays between different areas in different animals

System	species	area	N	Distance (mm)	Conduction time (ms)	Conduction velocity (m/s)	Refs
Cortico-cortical	Rabbit	S1- S2	48		(2.0 – 28.9) mean:11.0	(0.3 – 4.6) mean:1.3	Swadlow, 1990
	Rabbit	V1 –V2		2.5-7.0	(2-12) mean:~6.5	(0.23-5.74) mean: 0.64	(Swadlow and Weyand, 1981)
	Rat	V1-V2	12		(3.66 - 7.87) mean:5.69	(0.291-0.6) mean: 0.413	(Nowak et al., 1997)
	Rat	V2-V1	11		(4.1 – 8.9) mean:6.00	(0.27 – 0.476) mean:0.377	(Nowak et al., 1997)
	Cat	S1- S2	26		<2 and >30		Miller, 1975
	Cat	S1-S2			<2 *		(Manzoni et al., 1979)
	Cat	Area 17/18 – area 19			<2	(9.0 -21)	(Toyama et al., 1974)
	Cat	S1 – Motor cortex			87% 1-2.2 Longest 6.8		(Waters et al., 1982)
	Cat	S1- Motor cortex			0.6 -7.2 mean: 2.5		(Zarzecki et al., 1983)
	Cat	Motor cortex – S1			90% <2 10% 7-16ms		(Deschenes, 1977)
	Monkey	V1-V2			(1 – 2.5) mean:1.1		(Girard et al., 2001)

System	species	area	N	Distance (mm)	Conduction time (ms)	Conduction velocity (m/s)	Refs
	Monkey	V2-V1			(0.25 – 4.5) mean:1.25		(Girard et al., 2001)
	Monkey	MT-V1	106		(1.0-1.7) mean:1.3		(Movshon et al., 1996)
	Monkey	LIP-FEF	329	30	(0.5 – 8.0) mean:2.3		Ferraina et al., 2002
Cortico-thalamic (layer 6)	Rabbit	V1	124	17	(2.0 – 42.7) mean:14.3	(0.4 – 9.6)	Swadlow and Weyand, 1987
	Cat	V1	134	20	(2.5 – 45.0)	(0.4 - 8.0)	Ferster and Lindstrom, 1983
	Monkey	V1	35		(2.0-20.0) mean:9.5		Briggs and Usrey, 2009
Thalamo-cortical	Rabbit	visual	127	17	(0.6 – 3.1) mean:1.2	(5.5 – 28.0) mean: 14.8	Swadlow and Weyand, 1985
	Cat	visual	250	~ 20	(0.3 – 9.7) mean:0.9	(2.1- 67.0) mean: 22.2	Cleland et al., 1976
	Cat	LGN-V1	171		(0.5 – 1) mean:0.62		(Toyama et al., 1974)
Corpus callosum	Rabbit	visual	40	~18	(2.4 - 39.8) mean:16.5	(0.7 – 7.5)	Swadlow, 1974a
	Cat	visual	36		(1.3 - 15.0) mean:2.7		Innocenti, 1980
	Cat	Sense-motor	87	10~20	(2.0 - 32.0) mean:10.1	(1.0 - 10.0)	Miller, 1975
	Monkey	Visual	51	~ 50	(2.6– 18.0) mean:7.0	(3.0 –23.0) mean:7.0	Swadlow et al., 1978

*Defined conductional delay as short latency

Response delay

Another type of delay is the response delay, which is usually measured in the visual system. By flashing a visual stimulus, at different stages in the visual system, there are different event related response times. This response delay could be measured directly using an electrophysiology method: recording the ERP time in different areas. Nowak and Bullier have reviewed this time delay explicitly (Nowak and Bullier, 1997).

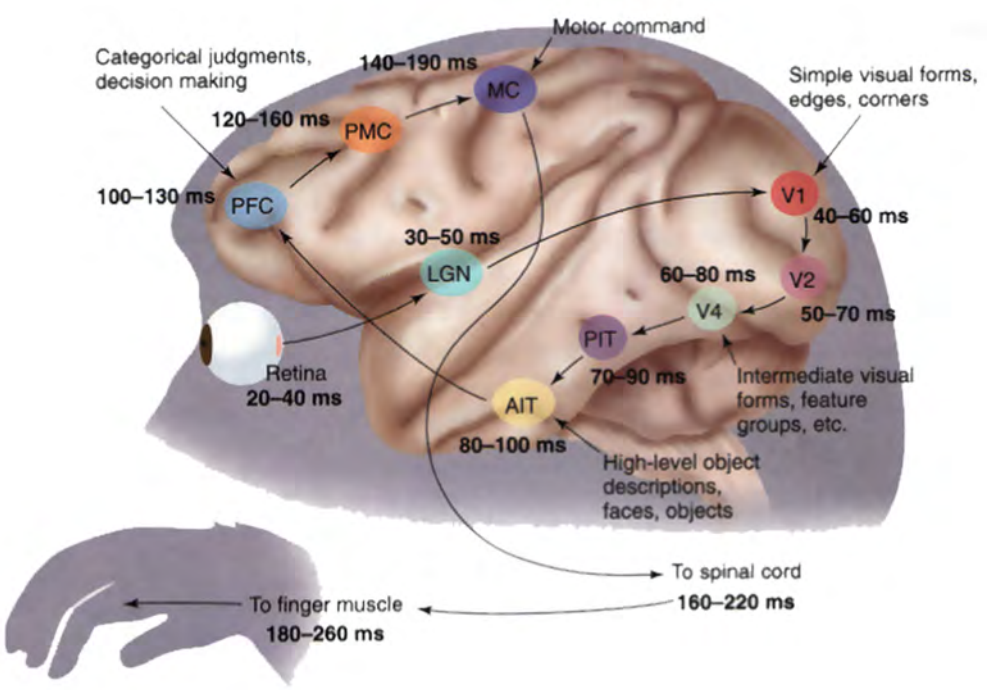


Figure 1-34 A time-delay map from Thorpe and Fabre-Thorpe. They suggested a 10 to 20 ms delay in every stage in the visual processing process. However, this conclusion was challenged by data obtained in Thorpe's later study. For example, they found a 24 ms mean onset latencies in FEF in epileptic patients. The new results suggested a much faster response time.

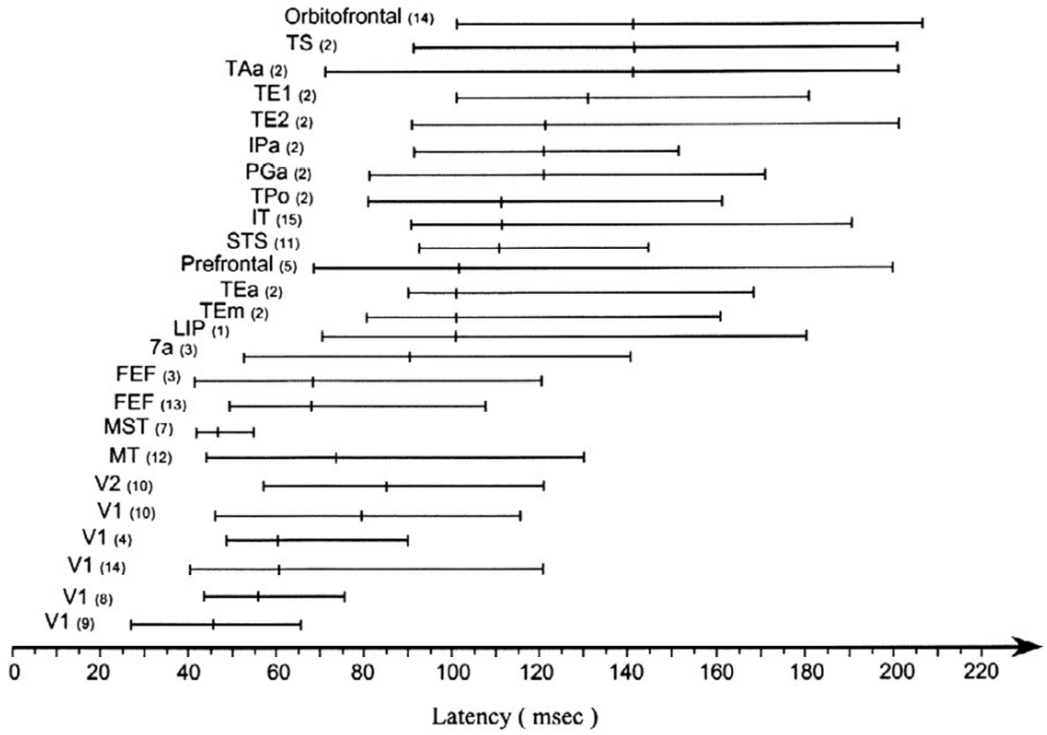


Figure 1-35 The onset response latency in different areas concluded by (Nowak and Bullier, 1997). The response time have a large variance.

Comparing with the axonal conduction delay, these delays have a much bigger value and variance. For example, the latency for V1 have a mean value of about 50 - 60 ms, but the latency for V2 is about 90 ms. This suggested a time delay between V1 and V2 of about 30 – 40 ms. Contrasted with the 1 – 2 ms axonal conduction delay, this delay is huge. Based on the response time to a flashed stimulus, Thorpe and Fabre-Thorpe showed a time delay map between different areas in monkey and concluded that it takes about 10 ms to 20 ms in every stage of signal processing (Thorpe and Fabre-Thorpe, 2001). Nowak and Bullier tried to explain this difference between the response latency and axonal conduction delay using the neuronal integration time: they showed previous evidence suggesting that for neurons in resting membrane

potential, it takes 5 – 12 ms for a neuron to integrate and fire for the optimally oriented stimuli; for the stimuli with close to optimal orientation, it takes 6 – 15 ms. However, they also showed evidence suggesting for a neuron that is close to the firing threshold, it takes only the EPSP rising time for the spike (which could be as short as 0.5 ms). They concluded that coincidence detection and temporal summation is the key factor for the response delay. I think the difference between the axonal conduction delay and response delay is the key computation window for the brain.

Electrical activity of the brain

The first discovery of the electrical excitability of the cerebral cortex was in 1870 by Fritsch and Hitzig and confirmed by Ferrier and others (Mountcastle, 1995). The electrical activity of the brain was first discovered in 1874 by Richard Caton. In 1875, he reported his discovery of this electrical activity in the grey matter in animals and hypothesized the possible functional role of this electrical activity.

The Electric Currents of the Brain. By RICHARD CATON, M.D., Liverpool.—After a brief *résumé* of previous investigations, the author gave an account of his own experiments on the brains of the rabbit and the monkey. The following is a brief summary of the principal results. In every brain hitherto examined, the galvanometer has indicated the existence of electric currents. The external surface of the grey matter is usually positive in relation to the surface of a section through it. Feeble currents of varying direction pass through the multiplier when the electrodes are placed on two points of the external surface, or one electrode on the grey matter, and one on the surface of the skull. The electric currents of the grey matter appear to have a relation to its function. When any part of the grey matter is in a state of functional activity, its electric current usually exhibits negative variation. For example, on the areas shown by Dr. Ferrier to be related to rotation of the head and to mastication, negative variation of the current was observed to occur whenever those two acts respectively were performed. Impressions through the senses were found to influence the currents of certain areas; *e. g.*, the currents of that part of the rabbit's brain which Dr. Ferrier has shown to be related to movements of the eyelids, were found to be markedly influenced by stimulation of the opposite retina by light.

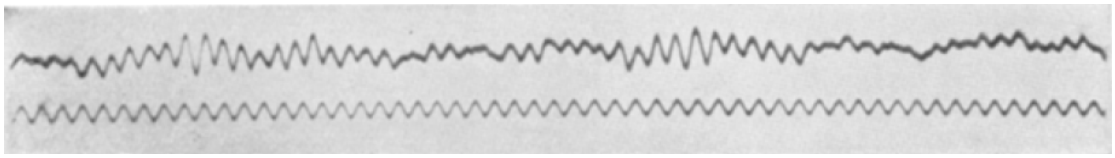


Figure 1-36 Top: Caton's discovery of electrical activity of the brain. From the proceedings of the forty-third annual meeting of the british medical association in 1875. Bottom: First published Electroencephalogram of a human by Hans Berger in 1929.

However, this discovery was not taken seriously for 55 years. Hans Berger followed the work by Caton and showed the electrical activity of humans in 1924. The neurological community was shocked by this discovery. Hans Berger's description is attractive because the usage of scalp recording technique (under the scalp) which was the first time that there is a method for studying the activity of the brain in waking, behaving human subjects (Mountcastle, 1995). Hans Berger also described the different waves or rhythms, such as the "Berger's wave" (~8 Hz – ~13 Hz). Adrian and Mathews confirmed that this discovery was not an artifact in 1934 and showed that these "alpha waves" were generated mainly in the occipital regions. These neural oscillations were then classified into Delta (~0.1 Hz – 3 Hz), Theta (~4 Hz - ~7 Hz) Alpha (~8 Hz - ~13 Hz), Beta (~14 Hz -~30 Hz) and Gamma (~30 Hz - ~100 Hz) frequency bands and were claimed to have different functional meanings.

Possible origins of the oscillations

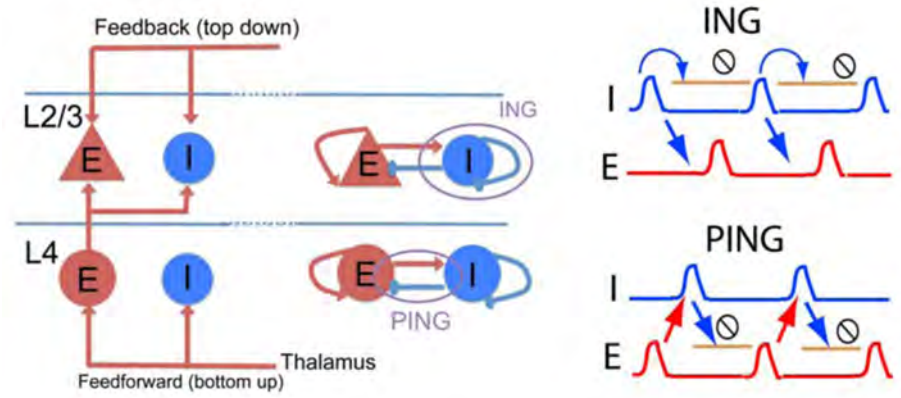


Figure 1-37 The ING model and PING model. Modified from (Tiesinga and Sejnowski, 2009)

The origins of the oscillations are from the rhythmic firing/spiking of a population of neurons and the interactions between the excitatory neurons and the inhibitory neurons creating these rhythmic activities. Even though we still do not know clearly about the detailed source of the oscillations, two models were created to try to explain the observed Gamma-band rhythmic activities: the interneuron Gamma (ING) and the pyramidal-interneuron Gamma (PING) (Tiesinga and Sejnowski, 2009).

The ING model states that the rhythmic activities are caused by the interactions within inhibitory interneurons and then affect the excitatory neurons: the inhibitory neurons inhibit themselves and generate a rhythmic synchronized spiking pattern. This firing pattern created a time window for the excitatory neurons to fire and thus the oscillations reflect the rhythm of the inhibitory neurons inhibiting themselves.

The PING model states that the rhythmic activities are caused by the interactions between the inhibitory interneurons and excitatory pyramidal neurons: the inhibitory neurons are driven by the excitatory neurons and begin to fire, while these activities inhibit the excitatory neurons in a circular manner. In this model, the rhythm of the network reflects the interaction between the excitatory and inhibitory neurons.

In both of the models, the oscillations are related to the inhibitory neurons. In the neocortex, Kätzel et al showed that the pyramidal neurons are connected to the interneurons within the same layer (Kätzel et al., 2011) (Figure 1-38). This suggested that different layers can have oscillations with different frequencies. From the definitions of feedforward and feedback, we know that feedforward and feedback connections rely on specific layers. Laminar recording studies showed that high-frequency oscillations are prominently generated in superficial layers and low-frequency oscillations in deep layers (Roopun et al., 2006; Maier et al., 2010; Buffalo et al., 2011). Since feedforward synapses are from superficial layers and connect to mostly middle layers, and feedback synapses are from deep layers and connect to mostly non-middle layers, the frequencies of the oscillations in feedforward and feedback connections can reflected by the oscillations in different layers. Thus, these results showed an oscillatory profile for feedforward and feedback: high frequency feedforward and low frequency feedback. Recent studies showed direct evidence for this notion: van Kerkoerle et al showed that, by inducing different frequency oscillations in different hierarchical areas with micro-stimulations and pharmacological method, the Gamma frequency oscillations propagate in the feedforward direction and Alpha frequency oscillations propagate in the feedback direction (van Kerkoerle et al., 2014).

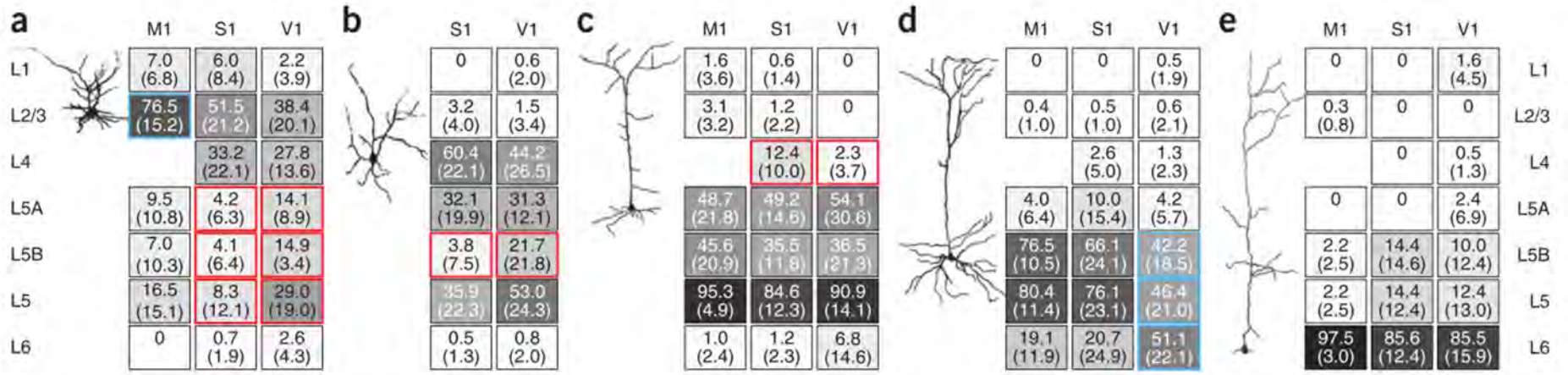


Figure 1-38 Average strength of inhibitory input from the indicated source layers (rows) to excitatory neurons located in L2/3 (a), L4 (b), L5A (c), L5B (d) and L6 (e). The data were from 30 neurons in M1, 54 neurons in S1 and 53 neurons in V1. The strength of a connection is expressed as the average percentage of inhibitory charge flow arising from identified inputs in a layer. L5 represents the sum of L5A and L5B. Values are represented numerically (s.d. in parentheses) and by the intensity of gray shading. The figure is modified from (Kätzel et al., 2011).

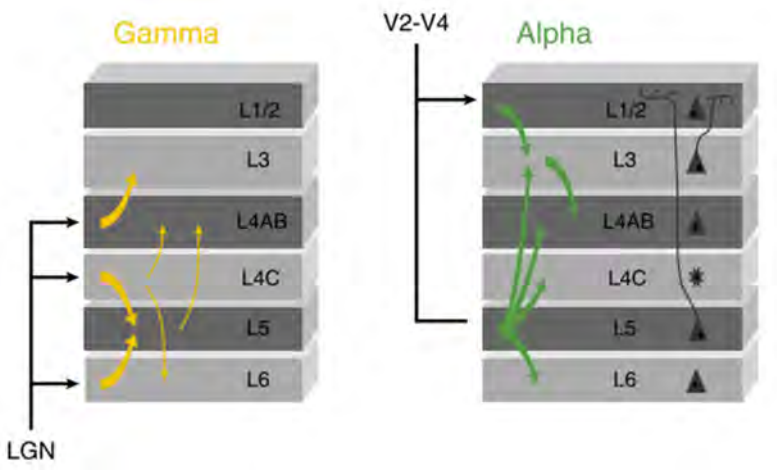


Figure 1-39 The top down and bottom up connections and their relationship with Gamma and Alpha band activity. From (van Kerkoerle et al., 2014).

Functional significance of ongoing oscillations

As the center of the machine controlling the action, the inner state of one's brain should have an effect on t behavior. The oscillations indeed have a big effect on the perception and behavior outcomes. Since the oscillations in each frequency have two properties: amplitude and phase, the investigations of the relationship between the behavior outcomes and amplitude/phase reveal the functional meaning of the different frequency oscillations.

The experiments usually use EEG since modern techniques can accurately record scalp activities with only 10 – 100 μV precision. The pre-stimulus oscillatory activities are usually used since the post-stimulus activities are driven to a large extent by stimulus-related activity (e.g. evoked potentials). These activities could hide the effect of the on-going oscillations and induce stimulus-related variability (e.g. eye-movement).

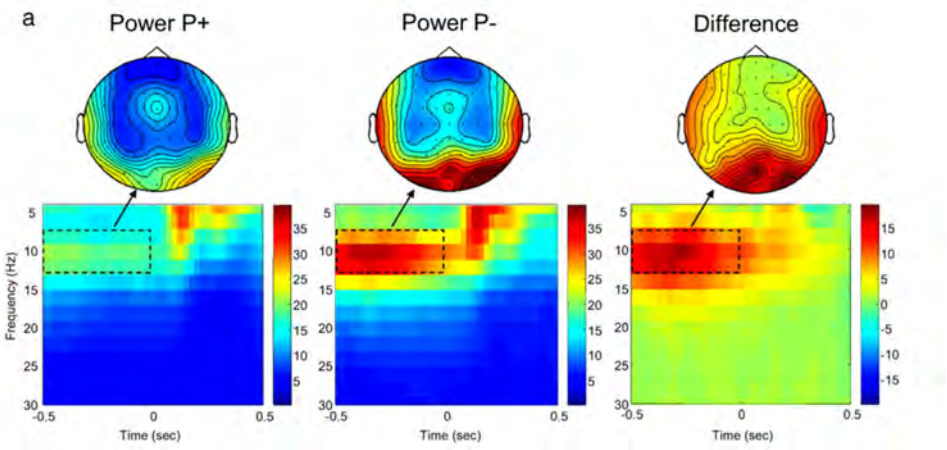


Figure 1-40 The pre-stimulus power difference in different perception. From (Hanslmayr et al., 2007)

One significant discovery on the relationship between the oscillatory amplitude and behavior is on the alpha frequency band. It was confirmed by several studies that there exists a negative relationship between the alpha amplitude and perceptual ability. One example is by Hanslmayr et al: in a discrimination task (discriminating the letters: p, q, b and d by button pressing), the subjects were asked to respond as fast as possible to the perception of a target stimulus. Results showed that, compared to the unperceivable conditions (P-), the perceivable conditions (P+) have less pre-stimulus alpha power (500ms until stimulus presentation) (Hanslmayr et al., 2007). Researchers concluded from this evidence that Alpha frequency oscillations have an inhibitory role in information processing.

Another very important discovery about the relationship between the oscillations and behavior is the phase-behavior relationship which was mainly discovered in our lab in 2009. Niko Busch and Rufin VanRullen published two papers on this: first they designed a visual detection task with a detection rate at ~50%. Subjects were asked to detect visual stimuli without moving their eyes

and report their perception while EEG was recorded. By using the measure of phase bifurcation index, they showed that pre-stimulus phase at ~7 Hz over fronto-central electrodes could influence the visual perception (the phase could decide ~16% perceptual performance) (Busch et al., 2009). Then, they confirmed that this effect only happened in a condition with attention (Busch and VanRullen, 2010) and contributed to the idea of a blinking spotlight of attention (VanRullen et al., 2007).

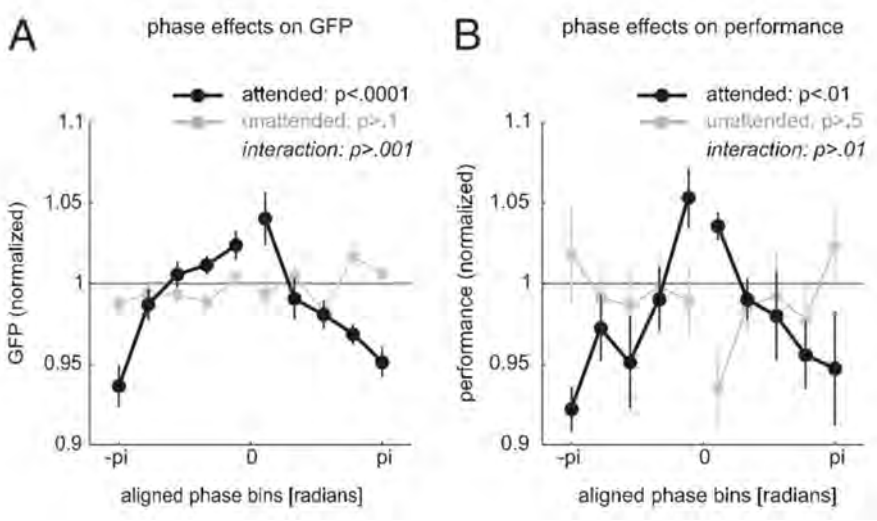


Figure 1-41 The relationship between the pre-stimulus phase and post-stimulus perception. From (Busch and VanRullen, 2010).

Canonical Neural Circuits

To sum up all the previous information about the brain, it is reasonable to build a canonical neural circuits model to further investigate the functional roles of different parts of the brain. From the beginning of neuroscience, researchers were beginning to search for canonical neural circuits and wanted to explain the brain using such circuits. Ramon Cajal was convinced that such canonical neural circuits exist. He claimed strongly that the neocortex was built of stereotyped circuits like those he had discovered in the other parts of the nervous system. However, he was not able to identify that canonical neural circuit (Douglas and Martin, 2007).

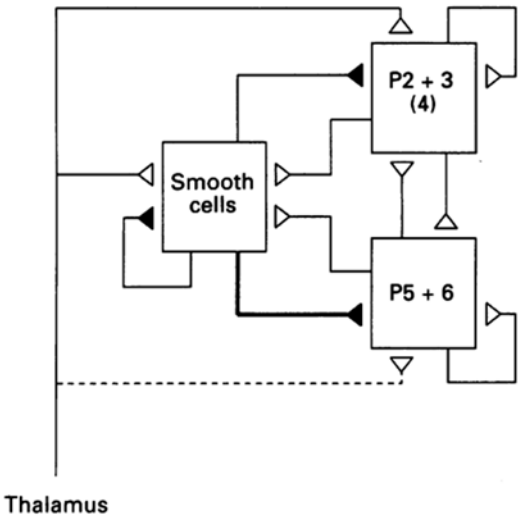


Figure 1-42 The canonical neural circuits proposed by (Douglas et al., 1989).

Douglas and Martin proposed their canonical neural circuits based on their recordings on cat. The circuit was first published in 1989 in a computational journal (Douglas et al., 1989) and then published in 1991 in a physiology journal (Douglas and Martin, 1991). In their circuit, they only described one stage of

processing (Thalamus to V1), in which the thalamus provides input to both smooth cells (inhibitory neurons) and pyramidal neurons in both layers 2, 3, 4 (mostly) and layers 5, 6. The superficial layers and the deep layers connected to each other and to the smooth cells. The smooth cells provided inhibitory input to superficial layers and deep layers, too. This model provided a certain understanding of the basic structure of the neural circuits, however, it lacked one very important property in the neocortex: the feedforward and feedback connections. They are one of the most important and well-studied property and provide information for the possible neural processing procedure.

Bastos et al. proposed a canonical neural circuit of predictive coding. They took advantage of a statistical description of the connections of different

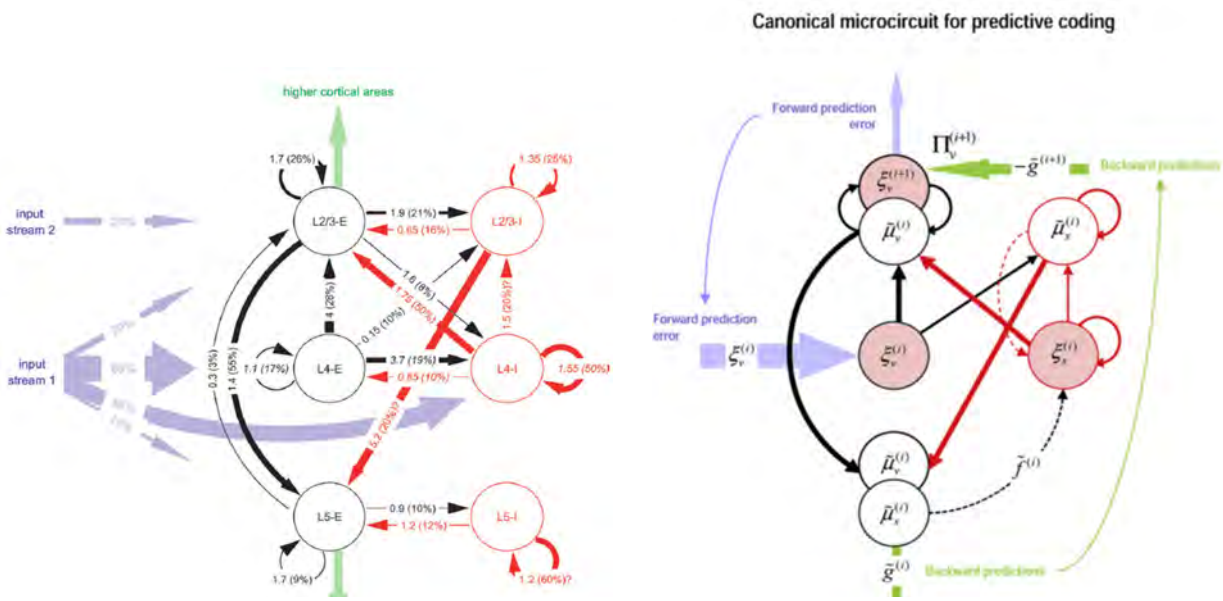


Figure 1-43 The proposed canonical neural circuits of predictive coding by (Bastos et al., 2012). The left is the original statistical description and the right is the proposed function of different groups of neurons.

layers. They simplified parts of the original description and added some connections which were not evident in the data (they used the dash line to represent them). Then, they assigned so called “error” population and so

called “representation” population to different parts of the neural elements. Even though this is a good attempt to try to fit the model with the empirical evidence, however, this circuit did not explain anything about the functional mechanism of predictive coding. For example, how is the predictive error generated? What is the functional significance of predictive coding?

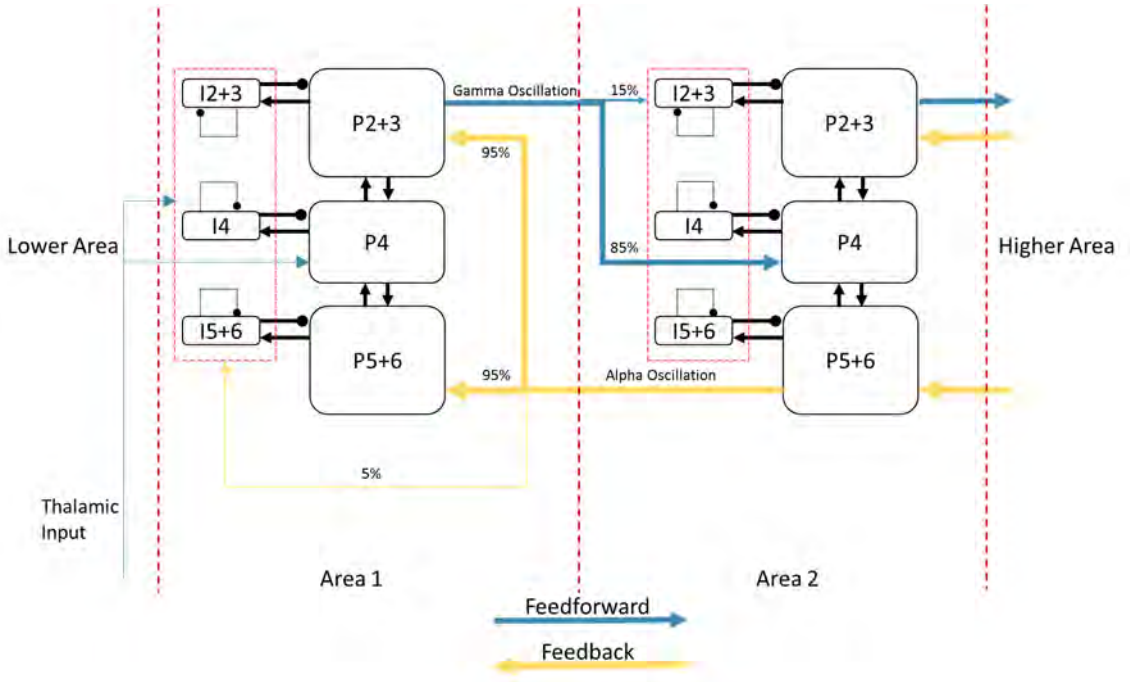


Figure 1-44 The proposed canonical neural circuit for distant areas (unilaminar). The circuit is based on the six-layer structure in the neocortex. We divided the six-layer into three parts: the surface layers (layer 2 and 3), middle layers (layer 4) and deep layers (layer 5 and 6). If the two areas are distant, feedforward connections (blue) projected only from the surface layer in the lower area with the Gamma oscillations generated within the local Pyramidal-inhibitory neuron loop. 85% of the feedforward connections target on the middle layer and 15% of them target on the inhibitory neurons. 95% of the feedback connection projected from the deep layers and targeted on the surface and deep layers (avoiding the middle layers) with the Alpha oscillations. The remaining 5% targeted on the inhibitory neurons. The feedforward and feedback connections are excitatory and the local inhibitory neurons provide the inhibition.

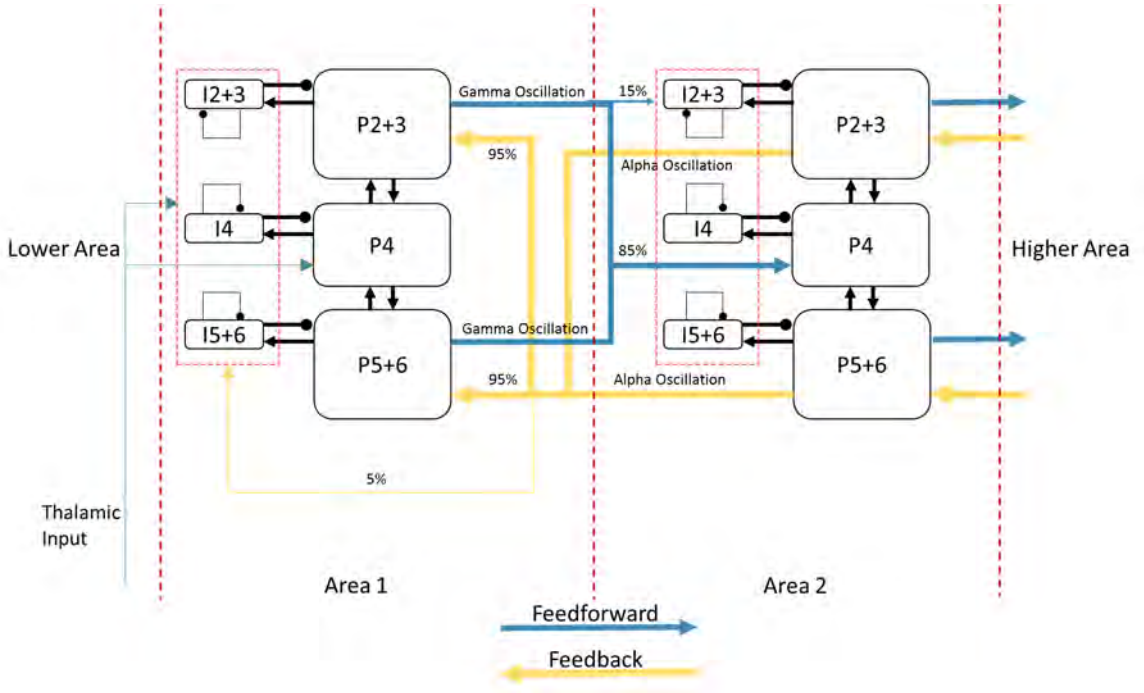


Figure 1-45 The proposed canonical neural circuit for neighboring areas (bilaminar). The structure is similar to the circuit showed before but with a bilaminar feedforward and feedback (projected from both surface layers and deep layers).

Based on the canonical neural circuits by Douglas and Martin and other recent evidence on the connections between areas, here, I propose one canonical neural circuit which includes the feedforward and feedback connections. These connections are from the classical research from Felleman and Van Essen (Felleman and Van Essen, 1991). Thus, there are two versions of the model: for neighboring areas, the connections are bilaminar; for distant areas, the connections are unilaminar. Since the anatomical evidence is strong and stable, any functional mechanism should be based on these basic structures. In the main text of this thesis, I propose one possible model for us to understand how such a simple and fixed structure could produce predictive coding, and further generate the significant functions of our brain.

Predictive coding

As usual, only more experiments, guided by the sort of insights provided by Rao and Ballard, will help unravel the complexities and multiple facets of information processing in the brain.

-Christof Koch and Tomaso Poggio

From efficient coding to predictive coding

Predictive coding is perceived as "the model" of the brain by many researchers. However, predictive coding was not born without any context. The idea of predictive coding comes from other ideas from more than 50 years ago. In 1948, Claude Shannon (1916-2001) published the classic paper "A Theory of Communication". This theory uses the amount of information to describe the world and gained such a success that changed the whole human society. Just few years after this discovery, scientists found out the similarity between the brain and a signal processing machine and had the idea to use the information theory to explain the brain. Fred Attneave (1919-1991) argued that many psychological facts of perception could be explained using the idea of information redundancy reduction by proposing several mind experiments (Attneave, 1954). In one of his mind experiments, he states:

…To begin, we give him an 80 X 50 sheet of graph paper, telling him that he is to guess whether each cell is white, black, or brown, starting in the lower left corner and proceeding across the first row, then across the second, and so on to the last cell in the upper right corner. Whenever he makes an error, he is allowed to guess a second and, if necessary, a third time until he is correct. He

keeps a record of the cells he has been over by filling in black and brown ones with pencil marks of appropriate color, leaving white ones blank. After a few errors at the beginning of the first row, he will discover that the next cell is "always" white, and predict accordingly. This prediction will be correct as far as Column 20, but on 21 it will be wrong. After a few more errors he will learn that "brown" is his best prediction, as in fact it is to the end of the row. Chances are good that the subject will assume the second row to be exactly like the first, in which case he will guess it with no errors; otherwise he may make an error or two at the beginning, or at the edge of the "table," as before. He is almost certain to be entirely correct on Row 3, and on subsequent rows through 20. On Row 21, however, it is equally certain that he will erroneously predict a transition from white to brown on Column 21, where the corner of the table is passed. (Attneave, 1954)

Attneave concluded from this mind experiment that the information redundancy exists in the graph paper. But in the same time, he pointed out one natural strategy to deal with redundancy: predicting the future and correcting the prediction with the errors.

In 1961, Horace Barlow (1921-) proposed the hypothesis of efficient coding. He proposed that the possible principles of sensory system included that “They recode sensory messages, extracting signals of high relative entropy from the highly redundant sensory input” (Barlow, 1961a). In 1972, Barlow proposed a more detailed version of this efficient coding theory, he stated: “The sensory system is organized to achieve as complete a representation of the sensory stimulus as possible with the minimum number of active neurons” (Barlow, 1972).

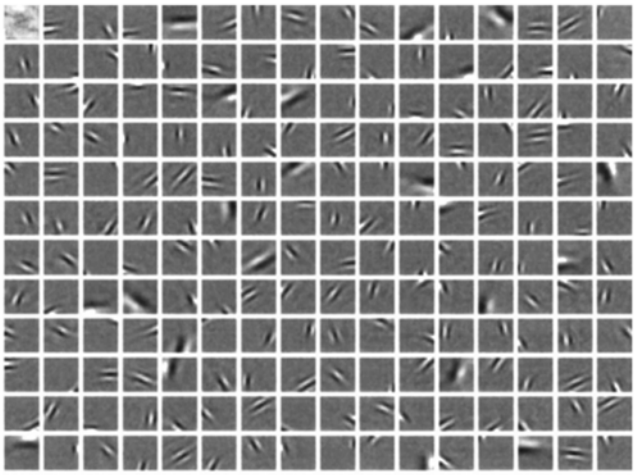


Figure 1-46 The learned basis functions (receptive field) using the sparse coding as prior. From (Olshausen, 1996).

In 1987, David Field provided evidence for this efficient coding idea. He found that the orientation and spatial-frequency tuning of mammalian simple cells suited well with the statistics of the natural images (Field, 1987). In 1996, Bruno Olshausen and David Field followed Barlow's idea and created sparse coding, which is literally to learn the basis functions (receptive field in the sense of neuroscience) based on the minimum number of active neurons. They found out the learned basis functions are just like the receptive fields in V1 (Olshausen, 1996) (see Figure 1-46). This algorithm gained a success in both the field of neuroscience and computer vision. In neuroscience, it could be one of the biggest discovery from the era since Hubel and Wiesel's discovery of the shape of the receptive fields. On the other hand, in computer vision, it is possible to solve many problems that traditional methods could not, as illustrated in Figure 1-47.

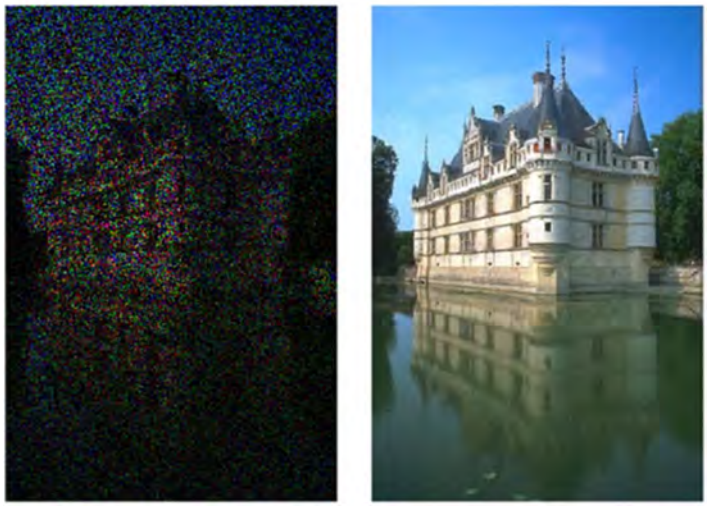


Figure 1-47 Image restoration achieved by sparse coding. The right image is restored using the information from left image and trained basis functions on natural images.

Rajesh Rao and Dana Ballard proposed the idea of predictive coding in 1999 which tried to provide a hypothesis for a fundamental brain mechanism. The study was motivated by the properties of extra-classical receptive-field such as the end-stopping, occlusion, perceptual grouping, illusory contours and etc. Then they argued the extra-classical receptive-fields are caused by the predictive coding of natural images.

In the proposed predictive coding model, there are three main components: the feedforward pathway, the feedback pathway and the predictive estimator. The feedback pathways carry predictions of neural activity at the lower level; the feedforward pathways carry residual errors between the predictions and actual neural activity. The predictive estimator uses the residual error to correct its current estimate of the input signal and generate the next prediction.

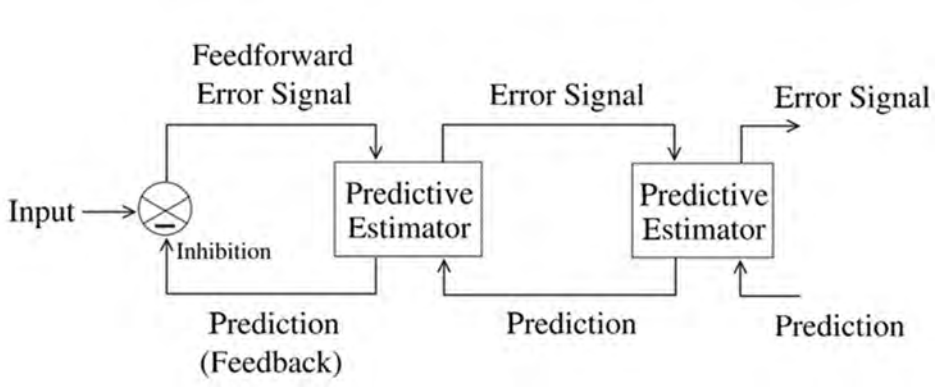


Figure 1-48 Proposed predictive coding model from (Rao and Ballard, 1999)

In their implementation, they used four kinds of neurons:

(1) Feedforward pathway neurons

(2) Feedback pathway neurons

(3) Error neurons, which stored the difference between the input signal and the feedback prediction

(4) Optimization neurons, which optimize the representation using gradient descent on the following cost function (E) with respect to the representation (r):

$$\frac{d\mathbf{r}}{dt} = -\frac{k_1}{2}\frac{\partial E}{\partial \mathbf{r}} = \underbrace{\frac{k_1}{\sigma^2}\mathbf{U}^{\mathrm{T}}\frac{\partial f^{\mathrm{T}}}{\partial \mathbf{x}}(\mathbf{I} - f(\mathbf{U}\mathbf{r}))}_{\text{Feedforward Representation}} + \underbrace{\frac{k_1}{\sigma_{td}^2}(\mathbf{r}^{td} - \mathbf{r})}_{\text{Feedback Representation}} - \frac{k_1}{2}g'(\mathbf{r})$$

which is to minimize the sum of the feedforward representation (1), difference between the feedback representation and input signal (3) and other control parameters.

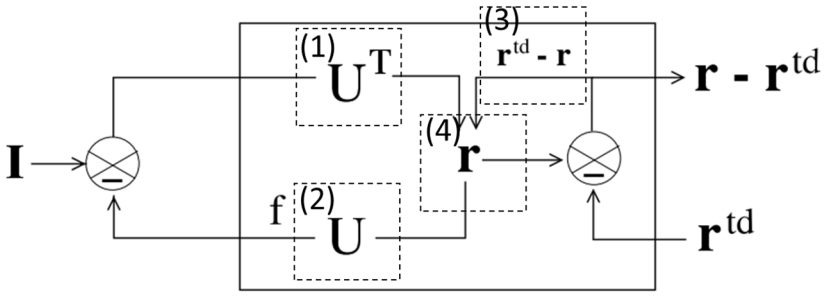


Figure 1-49 Different kinds of neurons in the predictive coding model

Their model could achieve a lot of functions of the extra-receptive field such as end stopping, pop-out texture, orientation contrast and etc.

After the publication of this predictive coding model, Christof Koch and Tomaso Poggio published a commentary on this model. They praised that:

Predictive coding is a general framework for interpreting information processing in complex natural and artificial systems, and many mechanisms may be seen in this light. (Koch and Poggio, 1999)

In 2005, Karl Friston developed his theory and tried to connect the predictive coding model with cortex laminar structure (Friston, 2005). However, his theory is hard to understand and applied mechanically the detailed implementations of Rao and Ballard's model, suggesting there are two groups of neurons inside the cortex: the representation neurons and error detecting neurons (which reflect type 4 and 3 neurons in the Rao and Ballard's model, respectively).

However, it is impossible to ask real neurons to do gradient descent. His misunderstanding had a huge and detrimental effect and many neuroscientists and psychologists tried very hard to find the "error neurons" and explained the excitatory feedback effect as "representation neurons".

In 2008, Michael Spratling contributed to the predictive coding framework and proposed the double-inhibition model to reconcile predictive coding with the neurophysiological and anatomical data showing that feedback is mainly excitatory.

In his model, instead of using direct inhibitory feedback to achieve the "error detecting" or "explaining away" effect, his model used a double-inhibition method: the higher area sends excitatory feedback to the representation neurons in the lower area; then these representation neurons send inhibitory input to the error neurons within one area; the error neurons receive excitatory input from a lower area and sends the excitatory output to the representation neurons. Thus, the simple error detecting neurons in Rao and Ballard changed from

$$\mathrm{r}\,(\mathrm{t}+1)\ \propto \mathrm{r}(\mathrm{t}) - \mathrm{r}_{\mathrm{td}}(\mathrm{t})$$

to

$$\mathrm{r}(t+1) \propto \mathrm{r}(t) + \mathrm{r}_{\mathrm{td}}(t) - 2\mathrm{r}_{\mathrm{td}}(t)$$

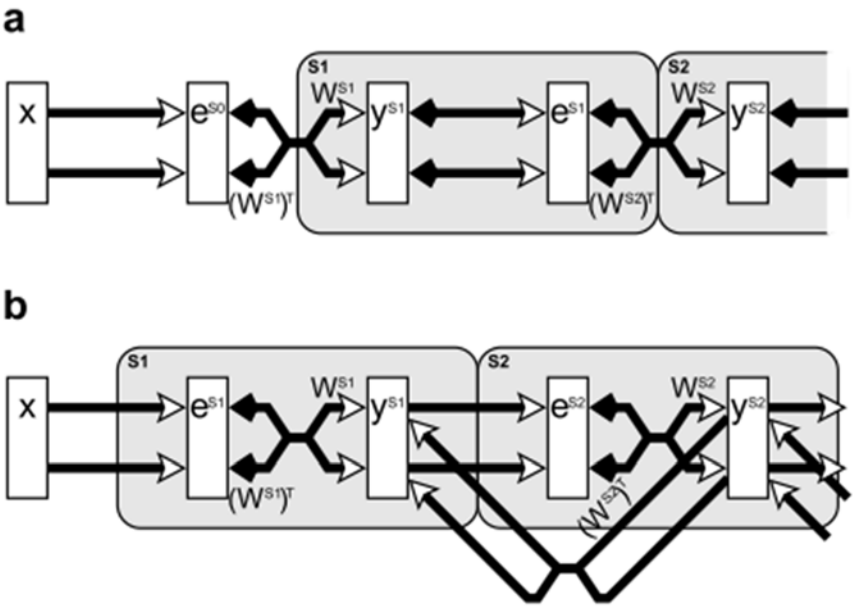


Figure 1-50 Proposed double inhibition model of predictive coding from (Spratling, 2008a) . a is the simplified Rao and Ballard's model, and b is proposed model.

In 2013, Andy Clark, a professor of philosophy, opened a discussion about predictive coding with the proposal that predictive coding is the future of cognitive science. Many researchers in psychology and neuroscience interested in predictive coding participated in this discussion (Clark, 2013). In the discussion, researchers mentioned the experimental approach to investigate predictive coding. In recent studies of predictive coding, the researches have three main topics:

(1) What is the effect of predictive coding?
(2) What is the relationship between predictive coding and attention?
(3) What is the relationship between predictive coding and oscillation?

I review some of the papers about these topics in the next section.

Empirical evidence of predictive coding

Effects of predictive coding

The first empirical evidence about predictive coding is from Murray et al in 2002. They used three fMRI experiments to try to prove that shape perception could reduce activity in V1 (Murray et al., 2002).

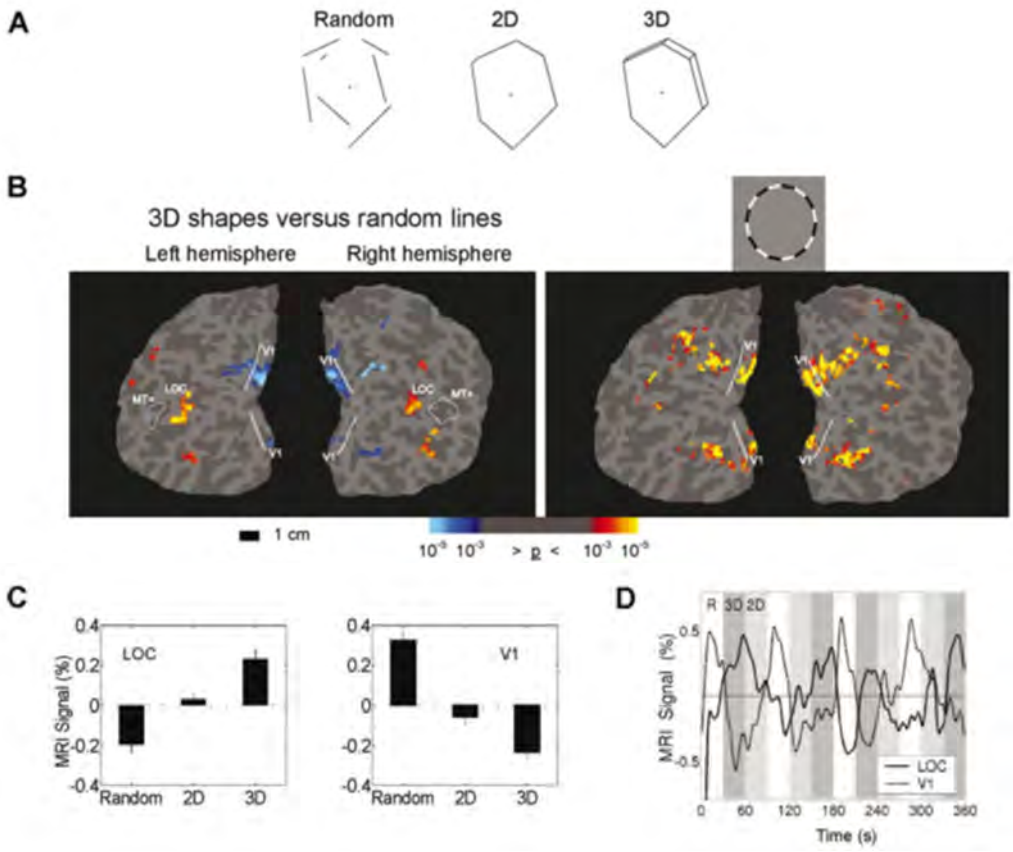


Figure 1-51 Predictive feedback inhibit primary visual cortex response. A is the stimuli that generate the predictive feedback. B is the different amount of activation in visual areas. C is the overall response in LOC and V1. D is the time-dependent response in LOC and V1. The results showed that 3D shape could activated the higher area (LOC) more and reduced activity in lower area (V1).

In the first experiment of their 3 experiments design, they used stimuli with random lines, 2D shape and 3D shape. They found that 3D shape stimuli produced more activity in LOC and less activity in V1 than the Random-line stimuli. Since the LOC is a higher area and should send predictive feedback to V1, Rao interpreted these activity decrease in V1 as the inhibitory effect caused by the predictive feedback. Murray et al did two other experiments (one using structure-from-motion and the other using Diamond motion) to confirm similar results of shape perception decreasing the activity in V1.

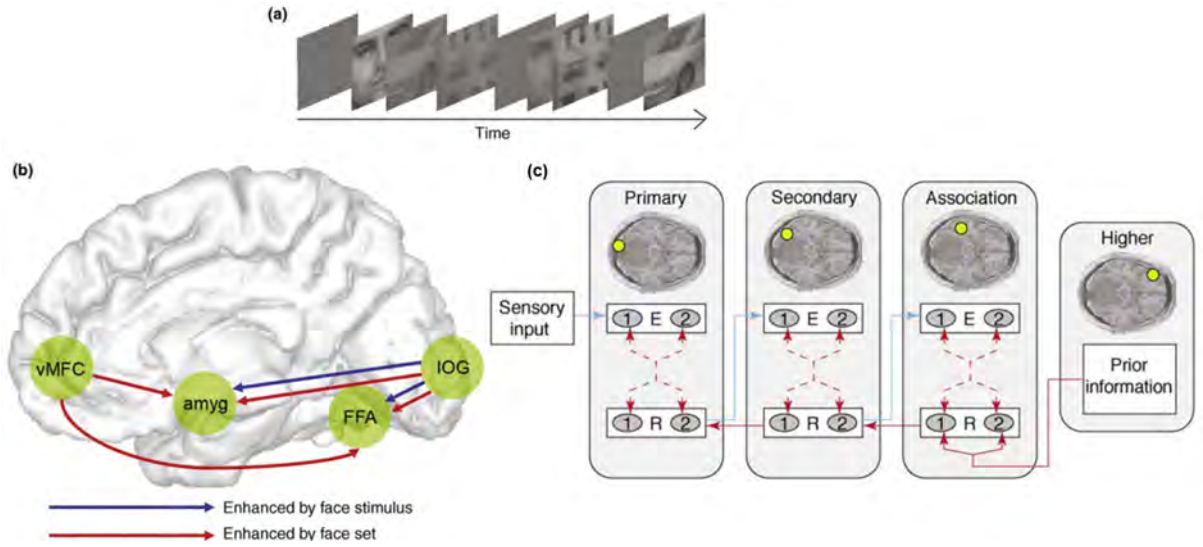


Figure 1-52 The connectivity increased by expectation. (a) subjects were asked to indicate whether the presented stimulus is a face or not (in 'face set' blocks) in a presentation of randomly intermixed degraded and masked images of faces, houses, and cars. (b) the task enhanced top-down connectivity from vMFC to amygdala and FFA, while both stimuli and task affected bottom-up connectivity from the IOG to the FFA and amygdala. (c)Proposed predictive coding's effect in visual perception. Modified from (Summerfield and Egner, 2009).

Summerfield et al. began to link expectation with predictive coding. They found more MFC activity and enhanced connections between the higher

area and lower area in the predicted condition (e.g. face stimuli in the block with face stimuli)(Summerfield et al., 2006). In this study, the face stimulus was presented for 100 ms, followed by a randomly selected mask (300 ms). The task was to discriminate face stimuli. In this study, they showed that in face stimuli block (with face expectation, or predictive feedback), the activity of face stimuli was higher than in the house stimuli block (without face expectation, or predictive feedback).

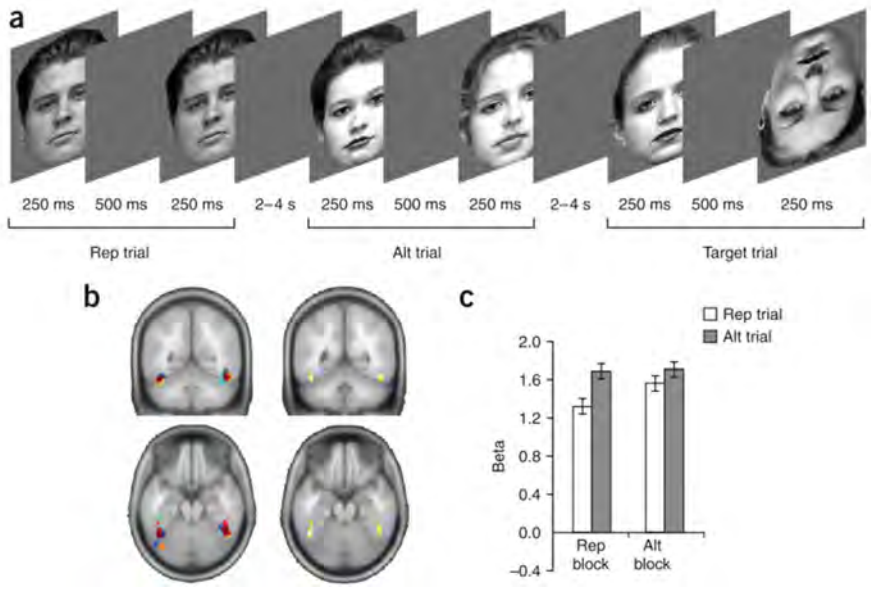


Figure 1-53 Repetition decreased face stimuli response in FFA. (Summerfield et al., 2008)

In 2008, Summerfield et al used fMRI to investigate the effect of repetition suppression. In the experiment, they compared the activity of the same face stimuli in the repetition trial and in the alteration trial. They found that the activity in FFA is lower in the repetition trial than in the alteration trial. Since repetition should send a predictive feedback to the lower area, they concluded that there is a relative reduction of the prediction error when the stimulus was expected, compared with an unexpected stimulus (Summerfield et al., 2008). In this experiment, the expectation cue (the first face) was shown

for 250 ms with a 500 ms gap with no stimuli on the screen and the target stimuli was shown for another 250 ms. The task was to detect upside down face stimuli. This finding is different from their 2006 result and the authors do not have a clear explanation for this difference, but we could take the different stimuli presentation time and tasks as parts of the reasons.

In 2009, Summerfield and Egner summed up the discoveries on expectation and claimed that the expectation is not the same as attention (Summerfield and Egner, 2009).

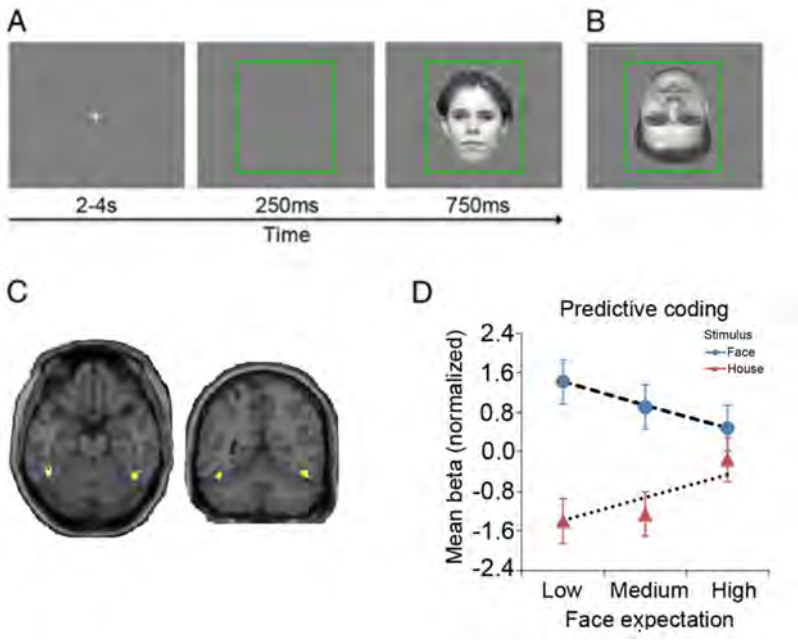


Figure 1-54 Expectation produced an activity pattern fitting the predictive coding model. Note that the main effect of different expectation in D is not significant.

In 2010, Egner et al showed that with different levels of expectation (low: 25%; medium: 50%; high: 75%; using different color box to indicate), the FFA activity fits better with the predictive coding model than feature-detection, baseline shift, multiplicative gain model (Egner et al., 2010).

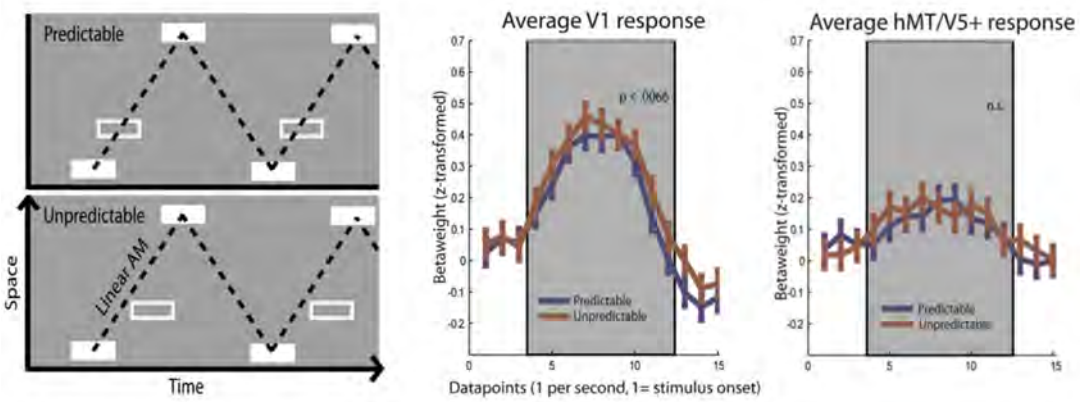


Figure 1-55 Prediction of a moving stimuli reduced the activity in V1 but not in MT.(Alink et al., 2010)

In 2010, Alink et al used a moving stimuli and found that the average V1 response is lower in the condition that the stimuli were in a predictable path (but the hMT response is basically the same), comparing to an unpredictable condition (Alink et al., 2010). This result suggested an inhibitory role for the predicted condition (with predictive feedback).

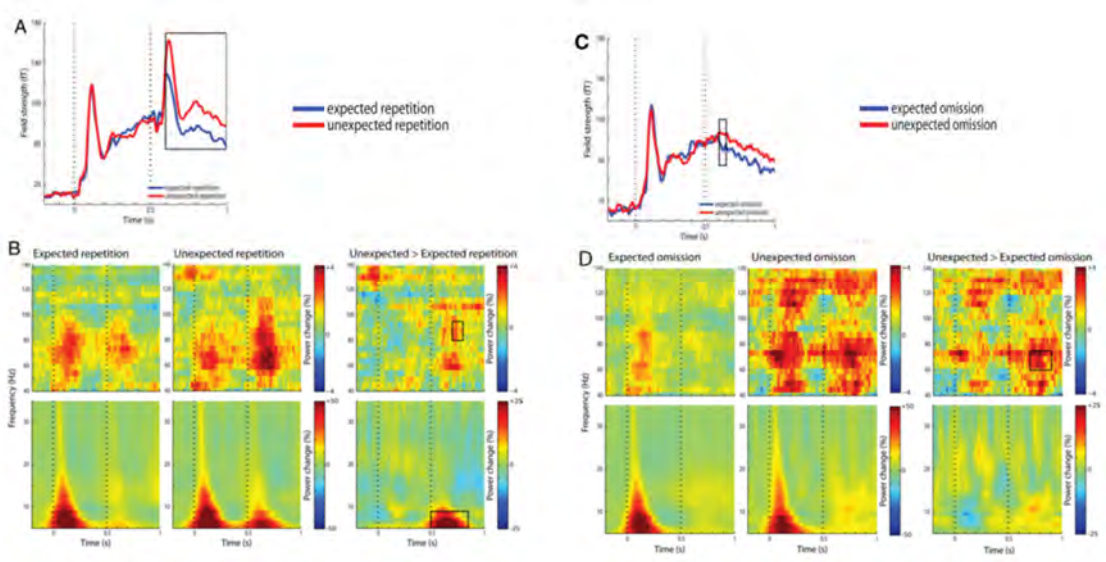


Figure 1-56 A,B: Expected repetition suppress Event Related Field (ERF) and gamma band power more than the unexpected repetition. C, D: Expected omission suppress ERF and gamma band power more than the unexpected omission. (Todorovic et al., 2011)

In 2011, Todorovic et al investigated the effect of expectation using auditory stimuli in blocks with expected/unexpected tone repetitions. By recording MEG, they found that repetition suppression was significantly larger for expected than unexpected repetitions in both ERF and Gamma-band activity. They concluded that predictive coding could help the repetition suppression.

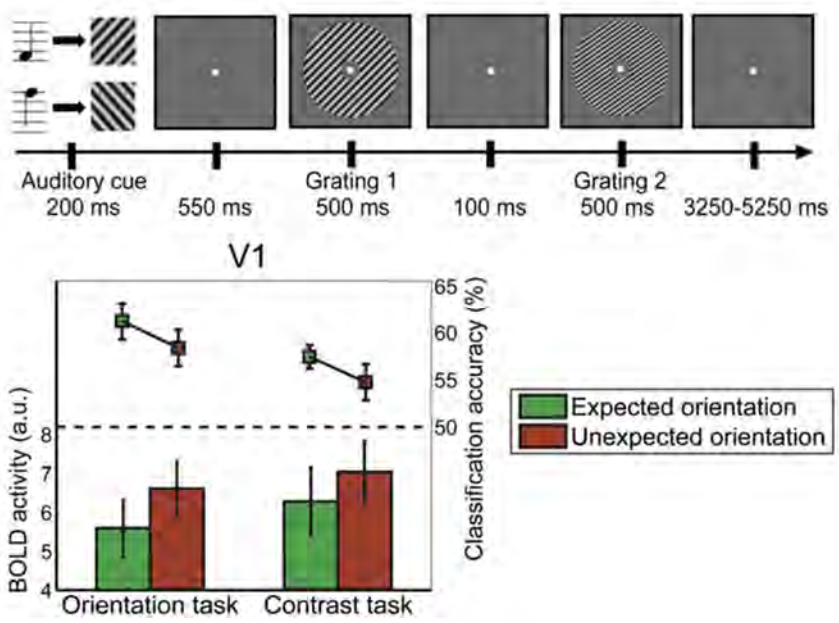


Figure 1-57 The expected orientation tilt in both orientation detection task and contrast task suppressed BOLD activity and increased MVPA classification accuracy. (Kok et al., 2012a)

In 2012, Kok et al used both orientation judgment task and contrast judgement task of two gratings in a sequence to investigate the effect of expectation. They found out, for gratings with an expected orientation, there were less activity in V1 than gratings with an unexpected orientation. In the same time, by using a MVPA method, the V1 orientation classifier accuracy was higher in the expected orientation condition than in the unexpected orientation condition. They concluded that expectation (predictive feedback) could lead to a better representation of orientation (Kok et al., 2012a). Furthermore, Rohenkohl et al showed a similar effect in temporal expectation (Rohenkohl et al., 2012). However, previous studies have showed that attention could bias the MVPA representation of the object (Reddy et al., 2009) which suggested the observed effect may not be caused by the predictive feedback but rather attention (or attention and predictive feedback are the same thing).

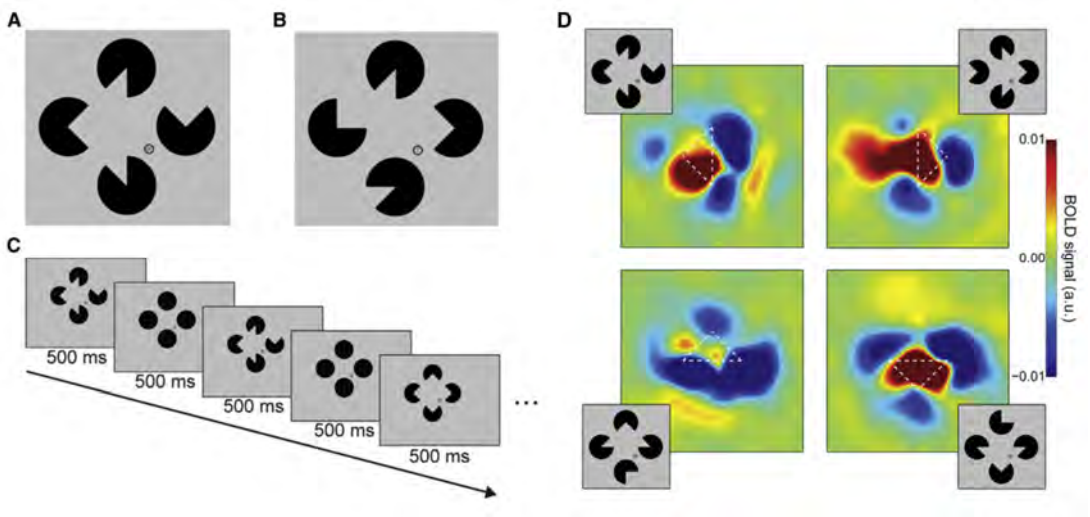


Figure 1-58 The contribution of the predictive feedback to different parts of V1 cortex. A is the used "Kanizsa" shape stimulus; B is the control stimulus; C is the experimental paradigm and D is the activity change in the retinotopic area in primary visual cortex. The results showed that the activity of the Pacman part reduced and the activity of the illusory part increased. (Kok and de Lange, 2014)

In 2014, Kok and de Lange used a "Kanizsa" stimulus to investigate the predictive feedback's contribution of different parts of the shape prediction. They found out that for a 500 ms illusory stimuli, the V1 cortex with a retinotopic position corresponding to the part inducing the illusion have an inhibitory effect from the feedback, the other part of V1 cortex have an excitatory effect.

To summary, we could conclude that:

(1) Predictive feedback could both reduce (Murray et al., 2002; Summerfield et al., 2008; Alink et al., 2010; Egner et al., 2010; Kok et al., 2012a; Kok and de Lange, 2014) and increase (Summerfield et al., 2006; Kok and de Lange, 2014) the lower area activity.

(2) Predictive coding could enhance the connectivity between higher/lower areas (Summerfield et al., 2006).
(3) Predictive coding could sharpen the representation (Kok et al., 2012a).

Predictive coding and attention

Another topic about predictive coding is its relationship with attention. Until 2009, the concepts of expectation/prediction and attention were not separated: the expectation was thought to be a part of attention. Even after Summerfield et al claimed that expectation and attention are different things, there was a debate about the relationship between expectation, attention and predictive coding. For example, Spratling created a predictive coding model which could be reconciled with the biased competition effect of attention (Spratling, 2008a). In the experimental literature, expectation and attention were defined as different operations (e.g. a cue in the beginning of the block as expectation and a cue just before the stimuli onset as attention).

In 2012, Kok et al found that attention could reverse the inhibitory effect of predictive feedback: in stimulus present condition, without attention, prediction reduced the activity of early visual cortex (significant in V1, N.S in V2 V3); with attention, prediction increased the activity of early visual cortex (significant in V1, V2, V3). In stimulus absent condition, attention increased activity for unpredicted omission. (Kok et al., 2012b)

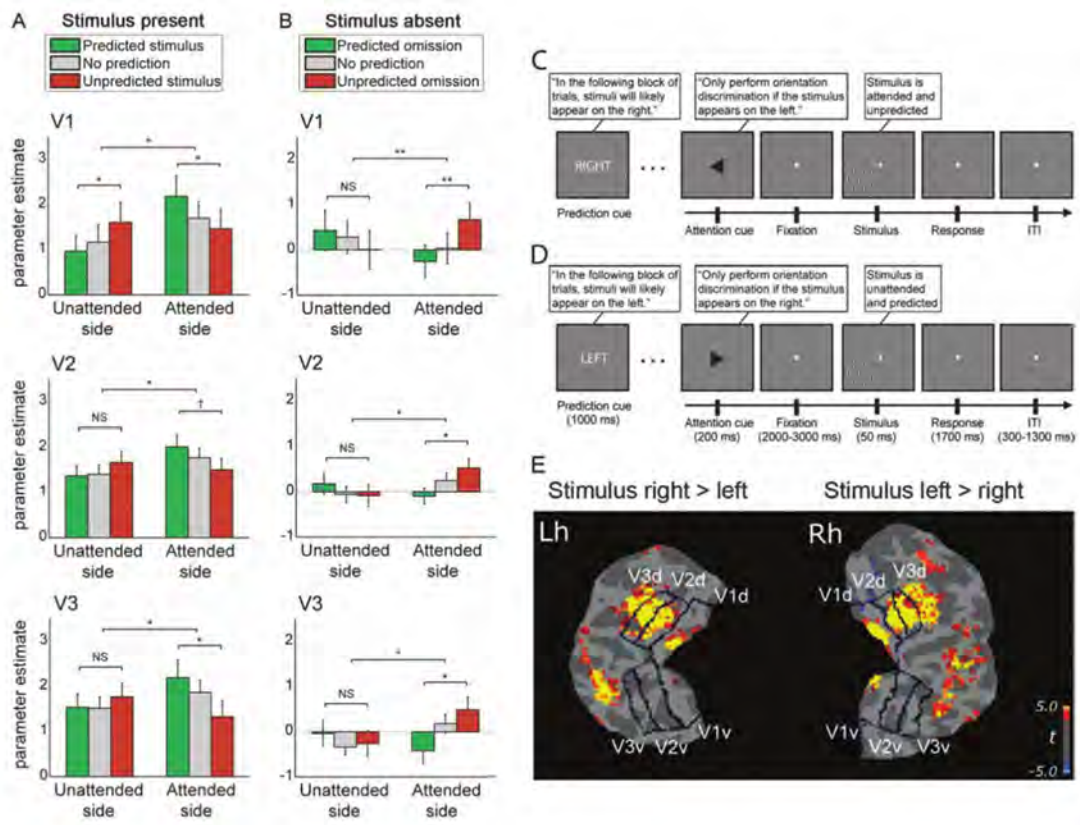


Figure 1-59 Attention reversed the silencing effect of predictive feedback. A and B are the activity change in different areas (V1, V2 and V3). C and D are the paradigms of attention and expectation in the experiment. E is the brain map and the general activity patterns of V1, V2 and V3.

In 2013, Jiang et al used a searchlight MVPA method on face and house/indoor stimuli with attention (detection task target) and expectation (audio cue before each trial), and found out that attention makes it easier to distinguish the expected/unexpected face stimuli in right FFA and expected/unexpected scene stimuli in the right PPA. (Jiang et al., 2013)

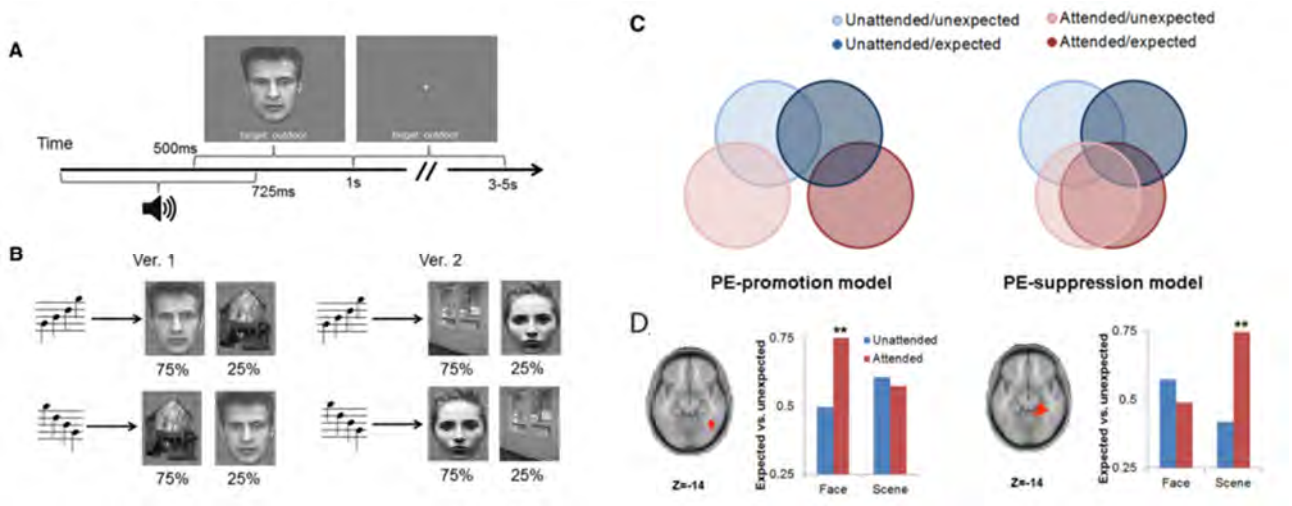


Figure 1-60 Attention makes the distinction between the expected and unexpected condition easier to identify. If attention promotes error signals, then the representations of unexpected stimuli more different from those of expected stimuli (C, left cluster), whereas the opposite would hold for attentional suppression of prediction errors (C, right cluster). The results showed that attention enhances the distinction between unexpected and expected stimuli (D). (Jiang et al., 2013)

There are few studies that focused on the relationship between attention and predictive coding. From these studies, there are no obvious conclusion. But at least, the evidence suggests predictive coding and attention are not two totally independent process.

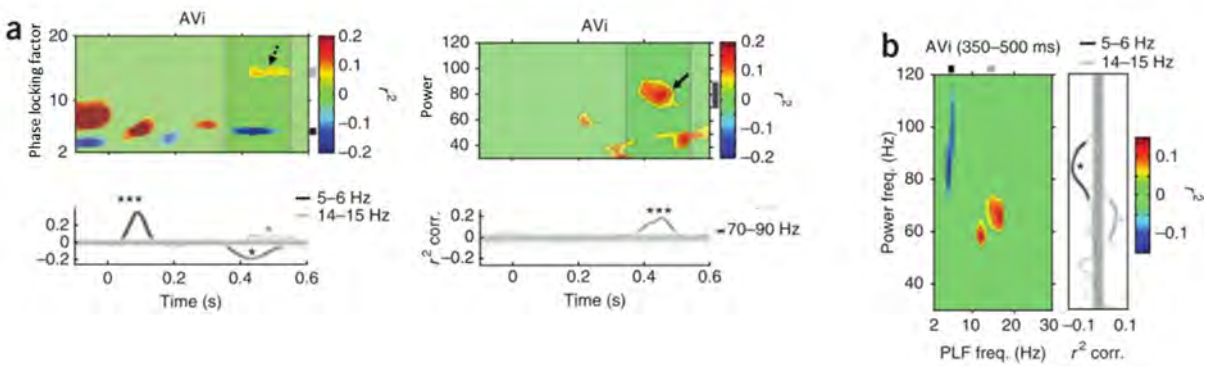


Figure 1-61 Audio-vision incongruent condition increased correlation between the ERF and phase locking in beta band and increased correlation between the ERF and power in Gamma band around 400ms after stimulus-onset. The beta-band phase locking also correlated with the gamma-band power. The author concluded from this results that predictive feedforward functions at gamma frequency and predictive feedback functions at beta frequency. Note that the beta frequency correlation increase is marginal significant and there is a more significant negative correlation between the theta frequency phase locking and ERF/gamma band power. (Arnal et al., 2011)

In 2011, Arnal et al used congruent/incongruent audiovisual speech stimuli to investigate the relationship between predictive coding and oscillations. They found that in the condition of incongruent audiovisual stimuli, there are marginally significant positive correlations between the phase locking factor in 15-16 Hz and ERF amplitude ($p<0.05$) and significant correlation between the power in 80-90 Hz and ERF amplitude ($p<0.001$) around 400ms after stimuli onset. There is also a significant negative correlation between the phase locking factor and ERF amplitude in 5-6 Hz around the same time. There is also a significate phase-power correlation in the period of 350 ms – 500 ms after stimuli onset. They did not mention the functional role of the observed theta frequency and all other effects observed in other time period. (Arnal et al., 2011)

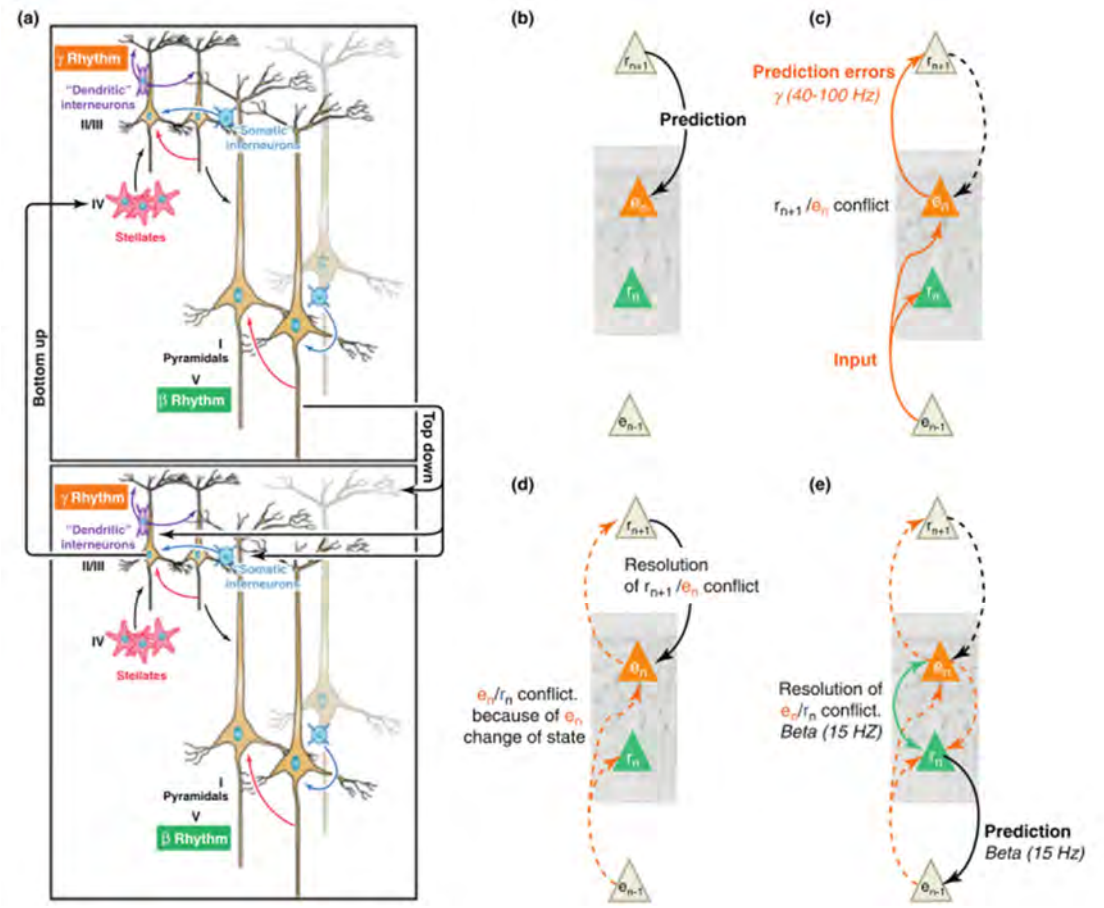


Figure 1-62 Proposed model of the relationship between the predictive coding and oscillations in different frequencies. (Arnal and Giraud, 2012)

In 2012, Arnal et al proposed a model for the relationship between predictive coding and oscillations. In the model, they proposed that the bottom-up (feedforward) pathway carries the predictive error, and top-down (feedback) pathway carries the prediction. Affected by Karl Friston, they also claimed there are representation and error units within one area of cortex. Then they assigned gamma band oscillation and beta band oscillation to the feedforward pathway and the beta band oscillation to the representation unit/error unit interaction.

The studies of predictive coding and oscillations are not as sound as other effects. There is room for further experiment to improve our understanding.

What's wrong and what's more?

For the state-of-the-arts studies about predictive coding, I have some concerns about the theory part and experimental part of predictive coding.

The concerns about the theory (modeling) part of the predictive coding are that:

(1) Is the predictive feedback always inhibitory? Even though from the Rao and Ballard, the predictive feedback was assumed to be inhibitory, the neurophysiological evidence suggested that feedback is dominantly excitatory. Spratling tried to reconcile the predictive coding model with the excitatory feedback, but the double-inhibition strategy seems to be not very economically sound since the system uses an excitatory interareal feedback and two inhibitory intrareal feedback steps to achieve the simple subtraction operation. Furthermore, there is no evidence for an accurate inhibitory intrareal connection since inhibitory neurons usually have a large effect on the overall area nearby.

(2) Does the distinction between the representation unit and error unit exist? Even though lots of the modeling work on predictive coding assumed these two units from Friston's interpretation of the Rao and Ballard's model, there is no direct evidence suggesting the existence of the two different groups of neurons. In their hypothesis, the superficial layers represent the error units and deep layers represent the representation

units. This should lead to two distinct types of receptive fields and firing patterns, however, there is no evidence for that.

About the experimental part of predictive coding, I think we could do more on the following points:

(1) Current studies mostly used the fMRI method which has a bad temporal resolution. The evidence about predictive coding usually showed a decreased activity (except some of the work by Summerfield et al and Kok et al) in lower area, however, the neurophysiological evidence mentioned before showed that feedback should have the excitatory role. Therefore, it worth to use methods with a better resolution (psychophysics or EEG/MEG) to measure predictive coding's effect on each time point, there may be a rich temporal profile.
(2) Another concern is about predictive coding and oscillations. As described before, there is room for improvement in the work about the relationship between predictive coding and oscillations. It would be interesting to know the precise oscillatory frequency that predictive coding performs on. It would be also interesting to know what is the properties of these oscillations: are they changing with the task or stimuli? Do they have any functional meaning in the framework of predictive coding?

In this thesis, my studies about predictive coding and they included both theoretical and experimental parts about predictive coding:

In my investigation of the theoretical part of predictive coding, I asked the question: "What is a better neuronal model for predictive coding under our current knowledge about the brain?"

In the investigation on empirical evidence of predictive coding, I asked two questions:

(1) What is the perceptual effect of predictive coding?
(2) What is the relationship between oscillations and predictive coding?

I think we could understand much more about the predictive coding and the brain after we answer these questions. Even though there is also room for improvement of my work, I do hope my work can help the researchers in this field to consider the predictive coding and observed brain properties as a whole. As a promising universal theory of brain, predictive coding theory requires this kind of thinking, and the reconciliation of predictive coding and neurophysiological and behavioral evidence could help us to understand predictive coding, and further to understand the working principles of our brain.

Chapter I

Predictive coding is a unifying framework for efficient coding in the nervous system. It uses the principle of eliminating the predictable neuronal responses and thereby permitting exclusive processing and transmission of unpredicted portions of the sensory input to promote an efficient way of coding.

It is obvious that the classical predictive coding model (Rao and Ballard, 1999) is not a neuronal model: it uses matrix instead of spiking neuronal network to represent the activity in neurons, and takes advantage of computational operations that are impossible in real neurons such as the "gradient descent". The most interesting thing in predictive coding is its underlying idea: using the feedback to achieve extra-classical receptive field effects by selectively inhibiting the predictable response.

To directly convert the classical predictive coding into a neuronal model, the simplest thing to do is to use the combinations of excitatory and inhibitory neurons to achieve a selective inhibitory feedback. For example, a neural network with a selective inhibitory neuron in the higher or lower area that directly inhibits the predictable response can generate this kind of selectivity. This is also the standard neuronal implementation of predictive coding: the feedback connections carry predictions of expected neural activity and the feedforward connections carry the residual activity between the predictions and initial lower area activity. To carry the residual, the feedforward connections are supposed to be excitatory, whereas to produce the residual the feedback connections are supposed to be inhibitory. To sum up, the

standard neuronal models of predictive coding hold that the different hierarchical levels interact by excitatory feedforward carrying residual activity and inhibitory feedback carrying predictions. Furthermore, we could interpret from this standard model that the neurons in one cortical area can be divided into two sub-populations, one coding for predictions/representations and one for prediction errors (Friston, 2005).

However, theories must follow facts. From the information we learned from the introduction part of this thesis, we know that physiological observations showed almost opposite evidence for the standard implementation of predictive coding:

(1) Most inter-areal feedback connections are excitatory and target excitatory neurons. Feedback usually project from excitatory neurons and targets on excitatory neurons. See more about the roles of feedforward and feedback connections in the introduction part of this thesis.

(2) Feedback usually exerts a divergent connection pattern. It has been shown by using tracers in the neurons that feedback connections target a much wider area in a lower area than feedforward connections do. See more about the convergence and divergence of the feedforward and feedback connections in the introduction part of this thesis.

Naturally, we asked the question: how could the brain implement the principles of predictive coding under such neuronal settings?

Correlated spike times create selective inhibition in a non-selective excitatory feedback network

Abstract

One of the most interesting contradictions in the study of neural networks relates to feedback inhibition. Specifically, feedback inhibition has been widely observed in the brain; however, most feedback connections and targeted neurons are excitatory. In addition, computational theories such as predictive coding suggest that such inhibition should be selective; however, neurophysiological observations indicate a divergent feedback pattern. Here, we propose a simple computational principle that essentially resolves these contradictions. We implement simple 2-layer hierarchical neural networks with non-selective excitatory feedback and demonstrate that it is possible to generate a selective inhibition effect by taking advantage of the spike time causality between lower and higher area neurons, together with a fundamental neuronal response property known as the "phase response curve". With computational modeling, we first show that lower area neurons are less responsive to feedback excitation (relative inhibition) when their spike times are correlated with those of active neurons in the higher area. This basic principle enables the feedback selectivity in a non-selective feedback network. Furthermore, we show that normalization in the lower area can turn the relative inhibition into absolute inhibition. The proposed computational principle provides a viable neuronal mechanism for efficient coding with a much more flexible spike-time based selectivity than traditional connection-weight based selectivity, and is supported by empirical evidence related to predictive coding.

Introduction

Neurons in the visual system follow a hierarchical structure: visual information flows from lower to higher cortical areas. Early studies demonstrated that neurons are excited by optimal stimuli in their classical receptive field (CRF) (Hubel and Wiesel, 1965, 1968), while stimuli in the receptive field surround (extra-classical receptive field, ERF) usually result in inhibition (Blakemore and Tobin, 1972; Nelson and Frost, 1978; Allman et al., 1985; Gilbert and Wiesel, 1990; Knierim and van Essen, 1992; DeAngelis et al., 1994; Levitt and Lund, 1997).

Researchers from both computational neuroscience and neurophysiology have proposed a unique idea to explain the ERF effect: feedback connections are the most likely source of surround suppression (Rao and Ballard, 1999; Angelucci et al., 2002; Angelucci and Bullier, 2003). In computational neuroscience, this idea is also related to predictive coding, which suggests that the inhibitory effect of feedback is exerted selectively on active neurons in the lower area whose response drove specific neurons in the higher area (predictable response). This idea can be extended to any two hierarchically connected areas (Summerfield and Egner, 2009) and is supported by substantial empirical evidence of an inhibitory feedback effect (Hupé et al., 1998; Murray et al., 2002; Summerfield et al., 2008; Alink et al., 2010; Egner et al., 2010; Kok et al., 2012; Schneider et al., 2014). However, such selective inhibitory feedback appears to contradict other classical neurophysiological observations in the neural system.

Firstly, most inter-areal feedback connections are excitatory and target excitatory neurons (Johnson and Burkhalter, 1996, 1997). Since only the excitatory neurons have long enough axons to travel across different areas, it is physically impossible for other types of neurons in one area to send

information to a different area. Furthermore, electron microscope studies have shown that feedback targeted neurons are also mostly excitatory (Johnson and Burkhalter, 1996, 1997). Secondly, feedback connections are rather divergent. Using retrograde and anterograde tracers, researchers have shown that feedback connections target a much wider area in a lower area than feedforward connections do (Ferrer et al., 1988; Henry et al., 1991; Salin and Bullier, 1995). Likewise, feedback connections have a much wider area of effect than horizontal connections; in addition, higher hierarchical order feedback is wider than lower hierarchical order feedback (e.g. the feedback effect from MT to V1 is wider than from V2 to V1) (Angelucci et al., 2002; Angelucci and Bullier, 2003). Thus, the evidence suggests a divergent/non-selective and excitatory feedback connection.

In this paper, we tried to address the contradiction between the observed non-selective excitatory feedback connections and the selective inhibitory feedback effect required by theory by proposing a computational principle of spike-time based selectivity. We tested simple hierarchical neural networks and demonstrated that correlated spike time can turn non-selective excitatory feedback into selective inhibition. If we define the predictable response as the lower area activities that driving the higher area neurons and the unpredictable response as the lower area activities that not driving the higher area neurons, the proposed computational principle can inhibit the predicable response (relatively, comparing to the unpredictable response, or absolutely, comparing to without feedback). Thus, it is also a viable neuronal mechanism for predictive coding, the modern implementation of efficient coding.

Results

We propose a mechanism of spike-time based selectivity as follows:

1) A set A of active neurons in a lower area drive specific neurons B in a higher area, thus the spike times of A and B populations are causally related.

2) The higher area sends non-selective, divergent excitatory feedback to the lower area.

3) Although this will tend to drive activity uniformly across the lower area, those neurons that have fired recently (i.e., those that drove the higher area in the first place) will be less sensitive to that excitation. This lack of excitation is effectively a relative inhibition of the originally active cells, as required by theory.

Thus simple facts of spike timing could establish a selective modulation of responses despite the feedback itself having no selectivity.

In order to test this computational principle, we built several increasingly complex two-layer hierarchical neural networks:

1) A three-neuron network (two in the lower area, one in the higher area) to establish that the basic principle works, and to explore its dynamics as parameters are changed

2) A larger network (~100 cells in the lower area), to explore how the balance of excitation and inhibition in the lower area can transform relative inhibition into absolute inhibition.

Basic principle of spike-time based selectivity

To demonstrate the underlying principle of spike-time based selectivity, as shown in Figure 2-1A, we used the simplest possible non-selective excitatory feedback model architecture: one higher area excitatory neuron sending non-selective excitatory feedback to two lower area excitatory neurons. On the other hand, the feedforward connections are selective: the predictable neuron ($\mathrm{Ex}_{\mathrm{prd}+}$) drives the higher area neuron, while the unpredictable neuron ($\mathrm{Ex}_{\mathrm{prd}-}$) does not contribute to the higher area neuron's activity. The predictable and unpredictable neurons receive the same amount of external input. The axonal conduction delay between higher and lower areas was set according to experimental observations in monkey V1 and V2: 1.1ms for feedforward and 1.25ms for feedback (Girard et al., 2001).

The other fundamental neuronal property we took advantage of here is the phase/spike-time response curve (PRC). This curve represents the relationship between the injection time of an input spike or current (relative to the last output spike) and the next output spike's time advance for spiking neurons driven by a constant input. The spike time advance represents the change in the next output spike time caused by the additional injection, relative to the normal situation (without additional injection). Single-neuron recordings have shown that the PRC in a variety of neurons have a similar shape (Figure 2-1 B): a flat (or negative in type 2 PRC (Hansel et al., 1995)) spike time advance in the beginning of the curve (injection just after the neuron's last output spike), followed by an increase in spike time advance from the middle of the curve

(injection delayed after the neuron's last spike), with a decrease in spike time advance in the end usually due to the absolute time advance limitation (there is an output spike immediately after injection, but the possible time advance is short because the injection is already late in the cycle) (Reyes and Fetz, 1993; Galán et al., 2005a; Lengyel et al., 2005; Preyer and Butera, 2005; Goldberg et al., 2007; Tsubo et al., 2007; Kwag and Paulsen, 2009; Smeal et al., 2010). This curve reflects the fundamental time-related input/output properties of single neurons. It shows that the same input to a neuron will have different results dependent on input time, and that inputs just after the neuron's last spike have less effect than inputs at other time points.

In our simple network, by definition, predictable neurons drive higher area neurons and unpredictable neurons do not drive higher area neurons. Thus, there is a strong spike-time correlation between the predictable neurons and higher area neurons: higher area neurons tend to fire just after predictable neurons. If the higher area neurons then send non-selective feedback to both predictable neurons and unpredictable neurons, the feedback would arrive (on average) at different time points in their PRC: at the beginning of the curve for the predictable neurons (determined solely by the axonal conduction delay between different areas), but uniformly across the PRC for unpredictable neurons (Figure 2-1 B). The relationship between injection time and membrane potential, spiking activity and firing rates is shown in the supplemental materials (Figure 2-S1).

The proposed computational principle is based on these different spike-time advances for different feedback times (relative to the neuron's last output spike time). The predictable neurons receive feedback in a rather fixed time window (just after their last spike), thus, the feedback has very limited effect on their activity. On the other hand, the feedback time to unpredictable neurons

has no correlation with their last spike, thus it can increase their activity, on average by the same amount as the average spike time advance of the PRC. This difference in spike-time advance for the predicable and unpredictable neurons produces a relative inhibition in predictable neurons. Moreover, this difference (or selectivity) is solely dependent on the spike-time correlation, suggesting that the targets of the selectivity (the inhibited neurons) can be changed without changing any synaptic weight and the predictable neurons are always inhibited. Therefore, even with the exact same feedback, the different feedback time correlation for the predictable and unpredictable neurons could lead to a robust firing rate difference between them, and the difference is only decided by the functional roles of the neurons (predictable or unpredictable).

Taking advantage of the simple non-selective excitatory feedback model (Figure 2-1 A) and the phase/spike-time response curve (Figure 2-1 B), we verified the effect of spike-time correlation on firing rate using a neural network simulation. As shown in Figure 2-1 C, when there was no feedback from the higher area neuron (feedback was artificially turned off), we observed similar activity patterns in the two lower area cells, both in the spike raster plot (Figure 2-1 C upper panel, which shows the spike activity for 100 simulation repeats) and the averaged firing rate plot (Figure 2-1 C lower panel, which shows the average time-varying firing rate for 100 simulation repeats). However, when feedback was turned on (at 400ms), a robust spike frequency difference between the predictable neurons and unpredictable neurons emerged. The mean firing rate for predictable and unpredictable neurons exhibits a 10Hz difference, while the higher area neuron had a similar firing rate as the predictable neurons (Figure 2-1 D). These results show that correlated spike times between the predictable neurons and higher area neurons created a

robust selective inhibition for the predictable neurons (relative to the unpredictable neurons, or relative inhibition).

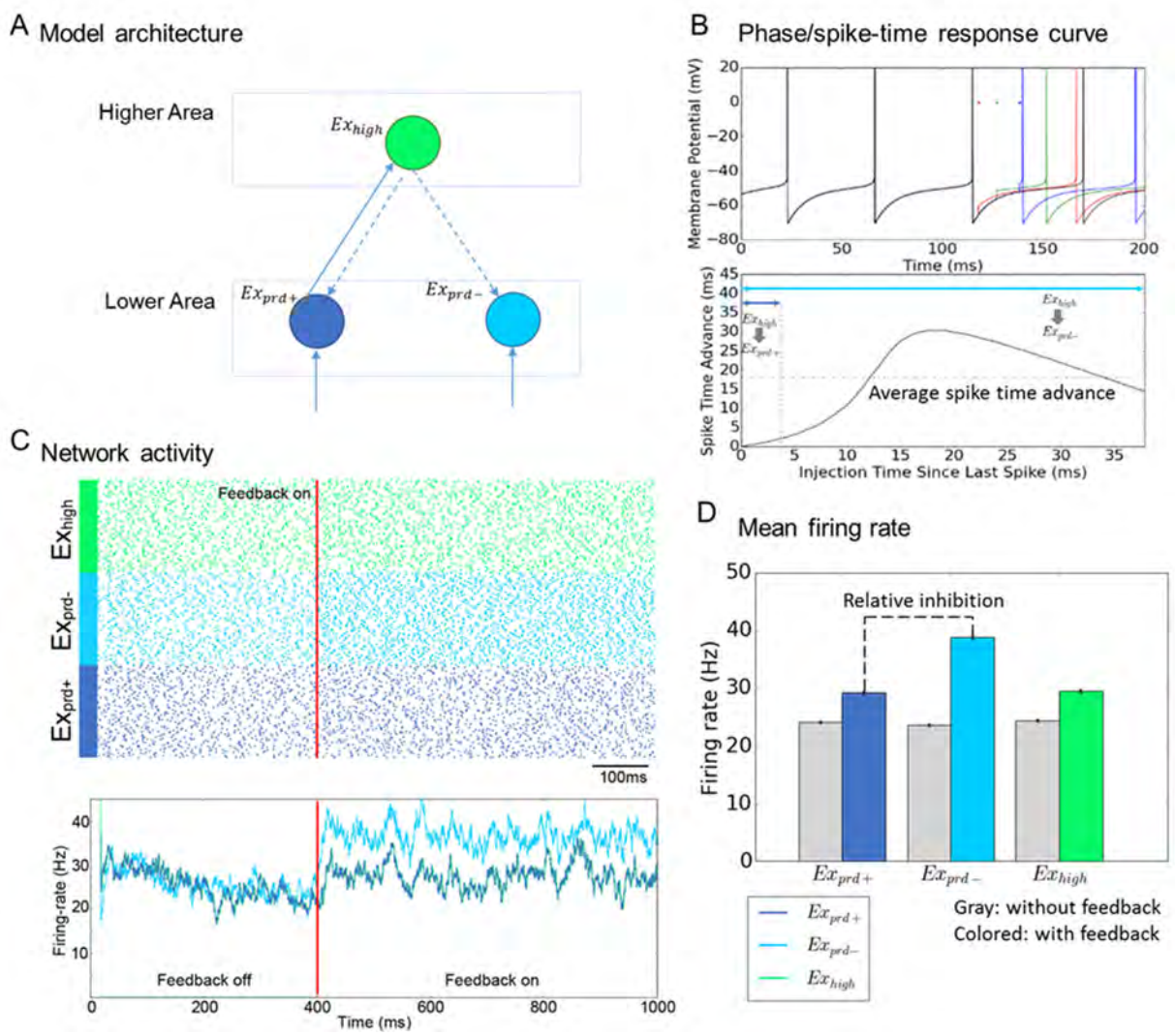


Figure 2-1 Basic principle of spike-time based selectivity. (A) The simplest non-selective excitatory feedback network. Ex_{high} neuron is the higher area neuron, Ex_{prd+} and Ex_{prd-} neurons are the predictable neurons and unpredictable neurons (neurons produce predicable/unpredictable response) in the lower area. Ex_{prd+} neuron drives Ex_{high} neuron and Ex_{high} neuron sends non-selective excitatory feedback to both Ex_{prd+} neuron and Ex_{prd-} neuron. The Ex_{prd+} neuron and Ex_{prd-} neuron receive same amount of Poisson spike input from the outside. The axonal conduction delay

between higher and lower areas was set to 1.1ms for feedforward and 1.25ms for feedback according to experimental observations. (B) The phase/spike-time response curve. Above, the same strength of injection with different injection time result different next spike time advance: the red, green and blue dot indicated the different injection time for the red, green and blue membrane potential trail. The black trail indicated the membrane potential without the injection. The relationship between the next spike time advance and injection time since last spike was plotted blow. They showed that at the injection just after last spike lead to less spike time advance than injection at other time points. Combining with the model's architecture, the feedback from Ex_{high} should land only at the beginning of Ex_{prd+} neuron's phase/spike-time response curve, while feedback from Ex_{high} could land at any time point of Ex_{prd-} neuron's phase/spike-time response curve. (C) Network activity of 100 times of simulations was showed. In the simulations, the feedback was artificially turned down in the first 400ms and then turned on. Both the spike raster plot above and the average time-varying firing rate showed a similar activity pattern for the Ex_{prd+} neuron and Ex_{prd-} neuron when the feedback were off. After turning on the feedback, more firing rate increase was observed in both the spike raster plot (more concentrated spikes) and time varying firing rate plot (higher firing rate) in Ex_{prd-} neuron than in Ex_{prd+} neuron. (D) Mean firing rate for different neurons in 100 times simulation (1000ms for each simulation) with feedback on and feedback off. Error bar indicated the standard derivation of different simulations.

Network Dynamics with different parameter settings

In our simple network, four key factors can affect the feedback and therefore affect the proposed computational principle: the feedback strength, the axonal conduction delays, the input noise, and the ability of predictable neurons to drive the higher area neurons. To investigate the effects of different

factors in the network, we measured the firing rate difference between the predictable neurons and unpredictable neurons with different parameter settings.

As shown in Figure 2-2 A, the feedback amplitude can modulate the spike-time based selectivity up and down: the effect of feedback strength on selectivity is not linear, but rather shows a peak at values around 10 mV, with low selectivity for both high and low strengths. We investigated the reasons underlying this result using phase/spike-time response curves with different feedback strength (Figure 2-S2). The results showed that the stronger feedback can lead to an increase in average spike time advance but the spike-time based selectivity also required a smaller axonal conduction delay between areas. Thus, the interactions between these two factors resulted in the observed relationship between the feedback strength and selectivity.

On the other hand, the axonal conduction delay between different areas showed a monotonic relationship with the spike-time based selectivity: the smaller the axonal conduction delay, the stronger the selectivity (Figure 2-2 B). The results also provided a reasonable time window for axonal conduction delay (selectivity emerged with less than 10ms total axonal conduction delay).

Furthermore, the proposed spike-time based selectivity showed very strong resistance to noise: the selectivity was still retained when the neurons received white noise with 1 nA variance (Figure 2-2 C; for the same simulated neuron, 1 nA input can generate 80 Hz spiking activity).

For the investigation on the relationship between the neuron's predictability (their ability to drive higher area neuron) and spike-time based selectivity, we used a different model architecture: the unpredictable neurons were

connected to the higher area neuron and could also contribute to its activity. The weight ratio between the predictable neurons to higher neurons and unpredictable neurons to higher neurons was adjusted to obtain different driving ability of predictable neurons (the higher the ratio, the stronger driving ability of the predictable neurons). Since the definition of the predictable and unpredictable neurons were based on their ability of driving the higher area neurons, an unpredictable neuron could easily turn to predictable neuron when the weight ratio is low. Thus, we used 100 neurons as unpredictable neurons and computed their average response as the response of the unpredictable neuron group (Figure 2-3 A). To generate differences in predictability, we had to change the feedforward weight for different groups of neurons in the lower area, however, this operation can potentially change the firing rate of higher area neurons and thus affect the feedback strength. Since we want to investigate the effect of predictability only, to avoid such change in feedback strength, we obtained a similar feedback (i.e. similar firing rate of higher area neurons and same feedback) with different feedforward weight ratio conditions while keeping the ratio of the feedforward input to the higher area. Results showed a monotonic relationship between the driving ability of the predictable neuron and the spike-time based selectivity: the stronger ability, the stronger the selectivity (Figure 2-3 B). These results suggest that the more predictable the neuron, the stronger inhibition it receives.

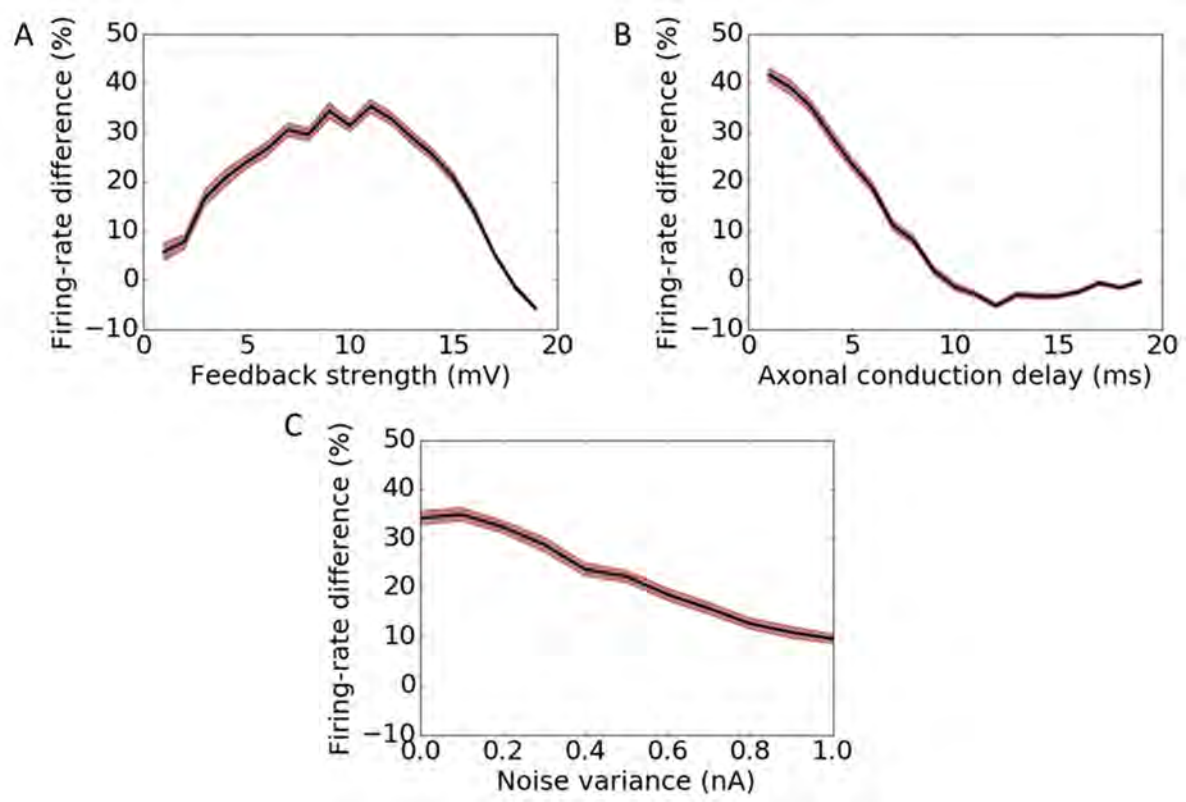


Figure 2-2 Firing rate difference between Ex_{prd+} and Ex_{prd-} with different parameter settings. (A) The relationship between the firing-rate difference and feedback strength. It showed that for a fixed axonal conduction delay between the higher and lower area, the feedback strength increase first increase the spike-time based selectivity and then decreased it. The possible reason behind the observed optimal feedback strength is that the increase of the feedback strength can increase the spike time advance, but in the same time, the selectivity is depended on the axonal conduction delay (as illustrated in Fig S2). (B) The relationship between the mean firing-rate difference between Ex_{prd+} and Ex_{prd-} and axonal conduction delay in 100 simulations. It showed that a smaller axonal conduction delay leads to a stronger spike-time based selectivity. (C) The relationship between the mean firing-rate difference and noise variance in single neurons in 100 simulations. Results showed that the spike-time based selectivity persisted with very high single neuron noise (In the same neuron, 1 nA input can generate about 80 Hz activity). Results suggested the proposed computational principle is very robust. Shaded area in all three plots represented the SEM of the firing rate difference across different simulations.

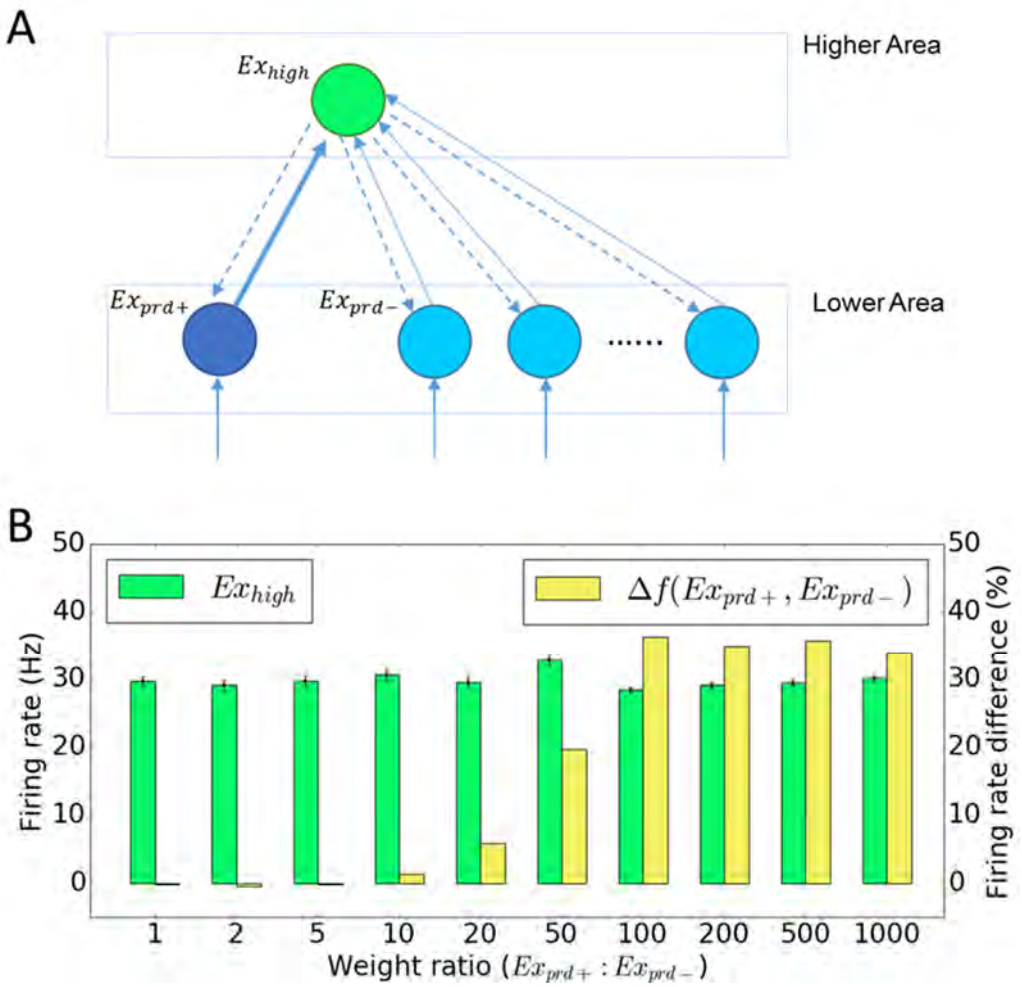


Figure 2-3 (A) The model architecture for the simulation with different Ex_{prd+} neurons driving ability. Instead of using the Ex_{prd+} as the only types of neurons that contributing to the higher area activity, both Ex_{prd+} and Ex_{prd-} neurons were set to connect with the higher area neurons with different weight. By adjust the weight ratio between the Ex_{prd+} to Ex_{high} and Ex_{prd-} to Ex_{high}, the driving ability of Ex_{prd+} neurons to Ex_{high} were modulated. Since it is possible that one single Ex_{prd-} neuron (not driving the higher area neurons) can be turned into Ex_{prd+} (driving the higher area neurons) when the weight ratio is low, Therefore more Ex_{prd-} neurons (100 neurons here) were used in the simulation and their mean response were used as the response of Ex_{prd-}. (E) The relationship between the weight ratio (Ex_{prd+} to Ex_{high} : Ex_{prd-} to Ex_{high}) and firing rate difference between Ex_{prd+} and Ex_{prd-} while keeping a similar Ex_{high} firing

rate and the same non-selectivity feedback strength. The green bar showed the mean firing rate of the Ex_{high} and the yellow bar showed the mean firing rate difference in 100 simulations. Error bar indicated the SEM across different simulations. Results showed that the stronger the driving ability of Ex_{prd+} is, the bigger the firing rate difference between the Ex_{prd+} and Ex_{prd-}.

The balance of excitation and inhibition converts relative inhibition into absolute inhibition

It has been shown that the inhibition generated in a cortical area is proportional to the total excitation (Vreeswijk and Sompolinsky, 1996; Anderson et al., 2000; Wehr and Zador, 2003; Zhang et al., 2003; Haider et al., 2006; Okun and Lampl, 2008; Atallah and Scanziani, 2009; Poo and Isaacson, 2009). For the proposed computational principle, the non-selective excitatory feedback from the higher area can increase the total excitation in the lower area, with the property that the predictable neurons receive less excitation and unpredictable neurons receive more. Such an increase in excitation should lead to an increase in inhibition, which could be able to convert the relative inhibition into absolute inhibition in certain conditions.

We tested this idea by adding a lower area inhibitory neuron into the model (Figure 2-5 A). This inhibitory neuron receives input from all lower area excitatory neurons and sends the inhibition to them with the same weight (In the literature, this is usually called "feedback inhibition"(Isaacson et al., 2011). Note that the term "feedback" here is different from the feedback in the proposed computational principle). Simulations showed that feedback from the higher area can increase the activity of the lower area inhibitory neuron (Figure 2-5 B, red bar). At the same time, it can generate spike-time based selectivity:

predictable neurons were relatively inhibited compared to unpredictable neurons (Figure 2-5 B, dark blue vs light blue bars). Furthermore, the predictable neurons were absolutely inhibited by the feedback: lower firing rates were observed with feedback than without (Figure 2-5 B, shaded dark blue bar vs the gray bar on its left side). On the other hand, the unpredictable neurons' activity was absolutely enhanced (Fig 2-5 B, light blue bar vs the gray bar on its left side). Thus, the observed results verified the idea that the balance of excitation and inhibition can turn the relative inhibition generated by the correlated spike-time into absolute inhibition.

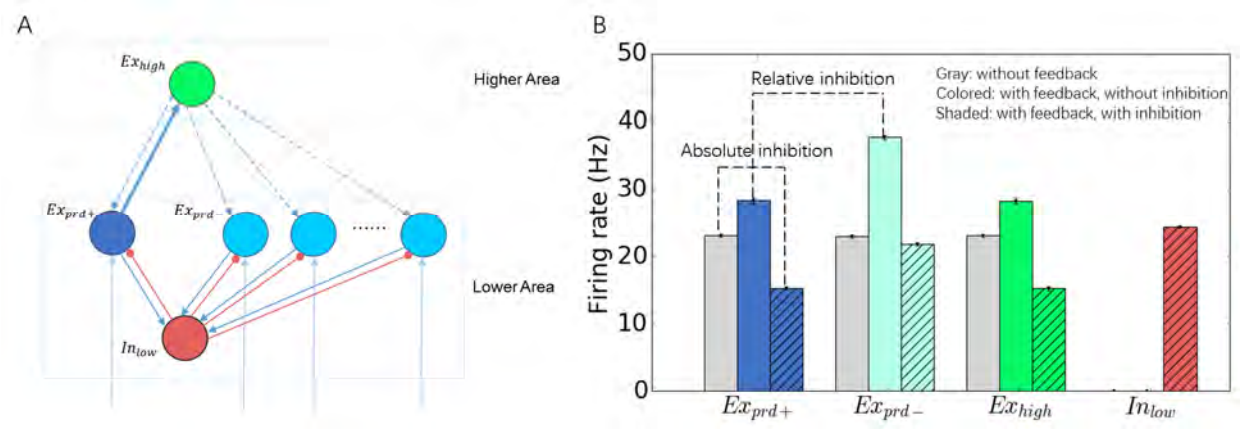


Figure 2-4 The balance of excitation and inhibition can convert the relative inhibition into absolute inhibition. (A) The model architecture is built upon the previous model (Figure 2-4 A) with an additional lower area inhibitory neuron ($\mathrm{In_{low}}$). The inhibitory neuron receives the same input from all lower area excitatory neurons (both $\mathrm{Ex_{prd+}}$ and $\mathrm{Ex_{prd-}}$) and sends the same inhibition to them. (B) The activity comparison between the without feedback (gray bars), with feedback but without inhibition (color bars) and with feedback and with inhibition (shaded bars). The results showed that, in a balanced excitation and inhibition network, feedback can produce the spike-time based selectivity (dark blue bar and light blue bar). In the same time, in the condition with inhibition, the increased activity in the lower area inhibitory neurons ($\mathrm{In_{low}}$, shaded red bar) can turn the relative inhibition (between the $\mathrm{Ex_{prd+}}$ and $\mathrm{Ex_{prd-}}$,

dark blue bar and light blue bar) into the absolute inhibition (between the $\mathrm{Ex}_{\mathrm{prd+}}$ *activities with and without feedback, shaded dark blue bar and the gray bar on its left).*

Discussion

One advantage of the proposed principle is that it fits well with our empirical understanding of the brain. Experimental evidence has shown that feedback connections are usually excitatory and divergent. In addition to the physical limitation that only excitatory neurons provide axons long enough to travel across different areas, these axons in feedback connections also usually target excitatory neurons. One study of the feedback connections between area LM and area 17 in rats showed that all of the observed feedback-targeted neurons in layer 1 are excitatory and only 2 out of 113 observed feedback connections targeted GABA-positive neurons in layer 2/3 (Johnson and Burkhalter, 1996). Similar results were obtained using a different labeling technique (Johnson and Burkhalter, 1997). Functionally, feedback has sometimes been found to be excitatory (Sandell and Schiller, 1982; Hupé et al., 1998; Wang et al., 2010). However, in contrast to the overwhelming evidence for excitatory feedback connections, evidence has also showed that feedback can often have an inhibitory effect (Hupé et al., 1998; Nassi et al., 2013; Zhang et al., 2014). For this inhibitory effect, the source of the inhibition was usually assumed to be local inhibitory neurons in the lower area (Zhang et al., 2014). Regarding convergence and divergence, using retrograde and anterograde tracers, researchers found more divergent feedback connections than feedforward connections (Henry et al., 1991; Salin and Bullier, 1995; Angelucci et al., 2002; Angelucci and Bullier, 2003) which suggests less selective feedback. Thus, the fundamental architecture in our proposed computational principle is founded on experimentally observed structures.

For the delays between different areas, although an average delay of 10 – 20ms for information to travel from one area to another has been reported (Nowak and Bullier, 1997; Thorpe and Fabre-Thorpe, 2001), axonal conduction delays are much smaller. In monkeys, the feedforward and feedback connections between V1 and V2 take median delays of 1.1ms and 1.25ms (Girard et al., 2001), respectively. The connections from MT to V1 only take 1.3ms (Movshon et al., 1996) and from LIP to FEF only take 2.3ms (Ferraina et al., 2002). In cats, from area 17/18 to area 19, it takes less than 2ms (Toyama et al., 1974). Similar values were observed in connections from S1 to S2 (Manzoni et al., 1979), from S1 to motor cortex (Waters et al., 1982; Zarzecki et al., 1983) and from motor cortex to S1 (Deschenes, 1977). These experimentally observed short axonal conduction delays make spike-time based selectivity possible.

Naturally, the proposed computational principle touches on the temporal coding *vs.* rate coding debate. Even though computational modeling studies have shown that temporal coding could be more accurate and carry more information (Van Rullen and Thorpe, 2001; VanRullen and Thorpe, 2002; Bohte, 2004; VanRullen et al., 2005), the rate coding scheme seems to be more intuitive: in different trials with the same input stimuli, the observed spike trains usually have a similar and reliable spiking rate (Adrian, 1926; Werner and Mountcastle, 1965; Tolhurst et al., 1983; Tolhurst, 1989; Britten and Shadlen, 1992; Tovee and Rolls, 1993; Petersen et al., 2000). However, on the other hand, the observed spike times are assumed to be too variable to support robust computation: the exact spike timing is random (which is usually modeled as a Poisson process) and the index of dispersion, spike counts' variance-to-mean ratios for the same stimuli, are near 1 (Tolhurst et al., 1983; Britten and Shadlen, 1993; Buracas et al., 1998; Shadlen and Newsome, 1998; McAdams and Maunsell, 1999). The proposed spike-time computational principle, however, relies on a more reliable kind of spike-time than the precise absolute spike

timing: spike-time correlation. The cortex is organized as a hierarchy (references), in which certain neurons in one area drive the activity of neurons in the next higher area. If these higher area neurons send direct divergent feedback to the lower area, there will be a spike-time correlation, and the robustness of this correlation is determined only by the variance of the conduction delays (which encompass the axonal conductional delay and neuronal integration time) and the robustness of the phase/spike-time response mechanism. The axonal conductional delays are very stable: the usual criteria for the delay jitter in experiments is less than 0.1ms (Girard et al., 2001). Furthermore, if the neurons in the feedforward pathway are doing coincidence detection rather than temporal integration (Softky and Koch, 1993; Roy and Alloway, 2001), the integration time may be negligible. Similarly, the phase/spike-time response curve is one of the fundamental properties of neurons, arising from the leaky nature of the cell membrane and has been robustly observed in experiments, and with similar shapes (Ermentrout, 1996; Netoff et al., 2004; Galán et al., 2005b; Lengyel et al., 2005; Tsubo et al., 2007; Kwag and Paulsen, 2009; Schultheiss et al., 2012). Thus, again, the computational mechanisms underlying the proposed spike-time based selectivity are well supported experimentally.

The proposed computational principle also provides a flexible coding scheme. On one side, the proposed principle provides a real time solution for efficient coding (Barlow, 1961). Traditionally, selectivity is provided by synaptic weights, which are ultimately set by synaptic plasticity, a long term process (significant change was observed after 20 min repeated stimulations in a spike-timing-dependent plasticity experiment (Bi and Poo, 1998)). These biological facts limit the flexibility of the neural network and it seems to be impossible to reduce the information redundancy (Barlow, 1961), or to inhibit predictable neurons (Rao and Ballard, 1999) in real time under this synaptic weight framework. The

proposed principle solves this problem using spike time correlation: the predictable neurons always get inhibited regardless of which specific neurons are involved, and the role of the neurons as unpredictable or predictable can evolve rapidly since the spike-time correlation is the only basis for selectivity. On the other side, the proposed principle provides a flexible definition for the higher area neurons. A higher area neuron needs to "know" which lower area neurons contributed to its activity in order to modulate them with feedback. For example, if a familiar face is represented in the higher area, the higher area needs to know which neurons in lower areas (e.g. simple cells in V1) contributed to the face perception in order to modulate them. However, many different lower level inputs can produce this face-specific response: we can recognize the same face under very different lighting conditions, points of view and distances, which correspond to very different groups of lower level neurons. In the synaptic weight framework, in order to send the appropriate feedback, for each lighting condition, point of view, and distance, one higher area face neuron needs to be created and set with a corresponding weight for each condition. Since the number of the possible scenarios is infinite, such arrangement seems implausible. In the proposed spike-time based selectivity framework, the problem is solved using a dynamic spike-time correlation instead of a fixed synaptic weight: only one higher area face neuron is needed and the feedback selectivity is automatically created.

The proposed model can also be a viable neuronal mechanism of predictive coding. Predictive coding (Rao and Ballard, 1999; Huang and Rao, 2011) is a framework for understanding redundancy reduction and a modern implementation of efficient coding theory (Barlow, 1961, 1972). In this hierarchical network framework, the feedback carries the prediction and explains away the predictable response in the lower area, while the feedforward only carries the residual errors between the predictions and

actual neural activity (Rao and Ballard, 1999). However, in the classical implementation of predictive coding, to selectively inhibit predictable response, it requires a complicated structure for feedback to mirror the synaptic weight patterns of the feedforward connections. In the proposed computational principle, the feedback achieves this function using the spike-timing correlation between the predictable response and higher area neuronal activity: predictable neurons are naturally inhibited in the model and activity in the remaining unpredictable neurons represents the error signal. Furthermore, the absolute inhibition in the proposed principle can explain the observed reduction in neural response in the lower area in the predictive coding literature (Murray et al., 2002; Summerfield et al., 2008; Alink et al., 2010; Egner et al., 2010; Kok et al., 2012). Using a "Kanizsa" illusion, it has been reported that the neurons in primary visual cortex corresponding to the illusory percept were inhibited by feedback, while the other neurons nearby were excited (Kok and de Lange, 2014). These observed activity patterns, with predictive feedback-induced excitation and inhibition for neurons with different roles, are compatible with our simulation (Figure 2-5 B). Therefore, the proposed computational principle can not only fit the predictive coding model, but also express activity patterns similar to the observed neural evidence.

To sum up, we proposed the computational principle of spike time based selectivity. Since the spike times of the higher area neuron are causally related to the spike times of certain neurons (predictable neurons, the neurons that drives higher area neuron) in the lower areas, robust temporal coding can be created using these spike times relationship. Especially, in a non-selective excitatory feedback network, the feedback can turn to be selective (relative inhibition for predictable neurons) because of the phase response curve. The balanced excitation and inhibition will turn the relative inhibition into absolute

inhibition (less activity for predictable neurons in condition with predictive feedback than without feedback). The proposed principle can help us to understand the redundancy reduction process of the brain and serve as a viable mechanism for predictive coding, the modern implementation of efficient coding.

Materials and Methods

Model Architectures and Neuron Types

In order to present the basic principle of spike-time based selectivity clearly and explore the dynamic in a larger and more complex environment, we adopt different model architectures in different simulations built on the same principle: a two-layer hierarchical neural network with non-selective excitatory feedback. Since it has been suggested that the horizontal connections are too slow and cover too small a part of the visual field to achieve the ERF related effect (Angelucci et al., 2002; Angelucci and Bullier, 2003), the excitatory neurons in the same area were set to be not connected to each other in our architectures. On the other hand, the inhibitory neurons act as the normalizing interneurons (Carandini and Heeger, 2011) and connected to (both sending signal to and receiving signal from) all the excitatory neurons in the same area.

In all model architectures, there are 4 types of neurons: higher area excitatory neuron ($\mathrm{Ex}_{\mathrm{high}}$), lower area predictable excitatory neuron ($\mathrm{Ex}_{\mathrm{prd}+}$), lower area unpredictable excitatory neuron ($\mathrm{Ex}_{\mathrm{prd}-}$), and lower area inhibitory neuron ($\mathrm{In}_{\mathrm{low}}$). The predicable neurons and unpredictable neurons are defined by their ability to drive the higher area neurons: predictable neurons are the

dominant driving force, while unpredictable neurons are not. Each model architecture is composed of some or all of the 4 types neurons.

Specifically, higher area excitatory neurons (i.e. located in the high-tier areas) receive feedforward input from the lower area neurons and send non-selective excitatory feedback to all lower area neurons. Given that lower area predictable neurons drive the higher area neurons as well, higher area neurons could obtain the representation of the lower area neurons and predict their response. Vice versa, the lower area unpredictable neurons do not drive the higher area neurons and cannot be predicted by the higher area neurons.

Pars	C	g_L	E_L	V_T	ΔT	τ	a
Values	281 pF	30 nS	-70.6 mV	-50.4 mV	1.5 mV	144 ms	4 nS

Table 0-1 Network parameter set

Neuronal Model and Synaptic Connections

To follow the observed neuronal response properties precisely, especially the phase response curve (PRC, or spike time response curve, STRC), we used a version of the conductance-based leaky integrate-and-fire model, specifically the adaptive exponential integrate-and-fire model (aEIF) (Brette and Gerstner,

2005), with random initial states in the simulations. In the model, the membrane potential obeys to the following equation:

$$C\frac{dV}{dt} = -g_L(V - E_L) + g_L\Delta_T e^{\frac{V-V_T}{\Delta_T}} - I_w + I_{syn} + \eta(t)$$

Where C is the membrane capacitance, g_L is the leak conductance, E_L is the resting potential, Δ_T is the slope factor, V_T is the threshold potential, $\mathrm{I_w}$ is an adaptation variable, I_{syn} is the synaptic current, and $\eta(t)$ is a Gaussian noise term. The adaptation variable $\mathrm{I_w}$ is defined by:

$$\tau\frac{dI_w}{dt} = a(V - E_L) - I_w$$

Where τ is the time constant and a represents the level of subthreshold adaptation. At spike time ($V > 20mV$), the membrane potential is turned back to the resting potential $\mathrm{E_L}$.

To demonstrate that the proposed principle does not depend on the absolute refractory period, we did not set any extra refractory term in the model. We used different parameters for the inhibitory neurons and excitatory neurons to fit the different neuronal characteristics, see Table 2-1. The parameters are modified from (Brette and Gerstner, 2005).

The external input to the network connect the lower area neurons using simulated Poisson input neurons where the synaptic current $I_{syn}(t)$ follows:

$$I_{syn}(t) = \sum_i I_i^{out}(t)$$

Where the I_i^{out} is the synaptic current from one single Poisson input neuron outside of the network. In both conditions, all lower area optimal neurons ($\mathrm{Ex}_{\mathrm{opt}+}$) receive the same outside input (same input current or connected to the same amount of Poisson input neurons with the same firing rates).

For the connections within the network, similar additive synaptic current equation was used, with an additional weight term:

$$\mathrm{I}_{\mathrm{syn}}(t) = \sum_i I_i^{in}(t) \cdot w$$

The conduction delays are considered when establishing the connections within the network. Since the proposed model is most likely to represent a neuronal mechanism in the early visual system, unless otherwise specified, we used the observed conduction delays between V1 and V2 in the simulations: 1.1 ms and 1.25 ms for feedforward and feedback connections, respectively (Girard et al., 2001).

Comparison Metrics

We used the traditional spike-rate based metric to measure spike-time based selectivity. Even though predictable ($\mathrm{Ex}_{\mathrm{prd}+}$) and unpredictable ($\mathrm{Ex}_{\mathrm{prd}-}$) neurons receive the same external input, and higher area neurons ($\mathrm{Ex}_{\mathrm{high}}$) send the same feedback to both types of neurons, we still expect that the predictable neurons are selectively inhibited. Thus, we used either overall firing rate to measure the inhibition over the total simulation duration or a sliding time-window firing rate to determine the dynamics of excitation and inhibition.

To evaluate the different effects on predictable neurons and unpredictable neurons, we used the percentage of the firing-rate difference (Δf) to measure the effect:

$$\Delta f(Ex_{prd+}, Ex_{prd-}) = \frac{freq(Ex_{prd-}) - freq(Ex_{prd+})}{freq(Ex_{prd+})}$$

Where the $freq(\mathrm{x})$ is the mean firing rate of neuron type x.

We investigated two types of inhibition in the simulations:

(1) Relative inhibition.

Since the only input difference between the lower area predictable and unpredictable neurons is the spike time correlation with the higher area neuron, we defined a lower firing rate in predictable neurons than in unpredictable neurons as relative inhibition.

(2) Absolute inhibition

Since the final goal is to investigate the contribution of the excitatory feedback to the predictable neurons, we defined a decreased response in predictable neurons with feedback than in predictable neurons without feedback as absolute inhibition.

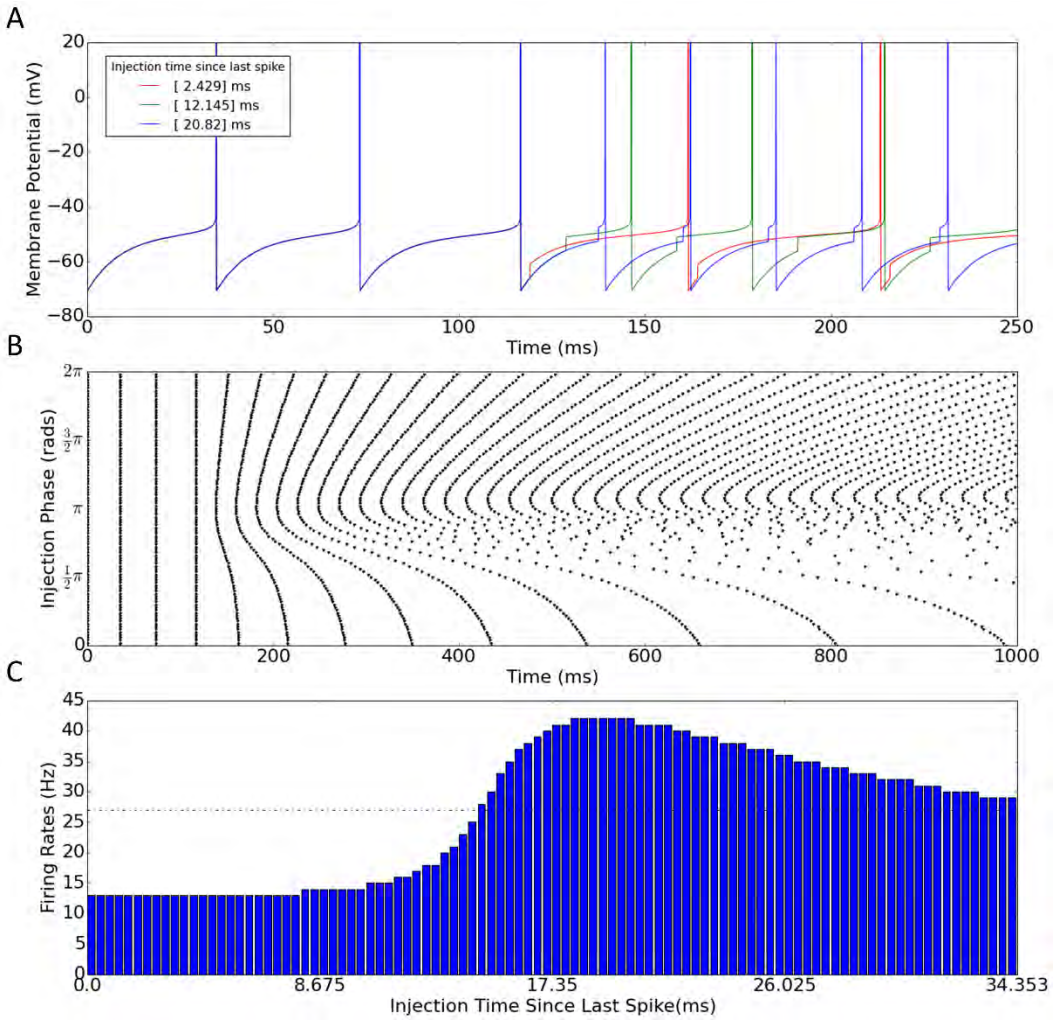


Figure 2-S 1 The different response with the same input in different phase/spike-time. (A) The relationship between the membrane potential and the different input injection time. It showed the same amount of initial increase in the membrane potential, but different next spike advancement. (B) The spike raster plot showed that the same input with different injection phase (relative to neuron's last spike) can lead to a difference spike time. (C) The relationship between the firing rate and the injection time since last spike.

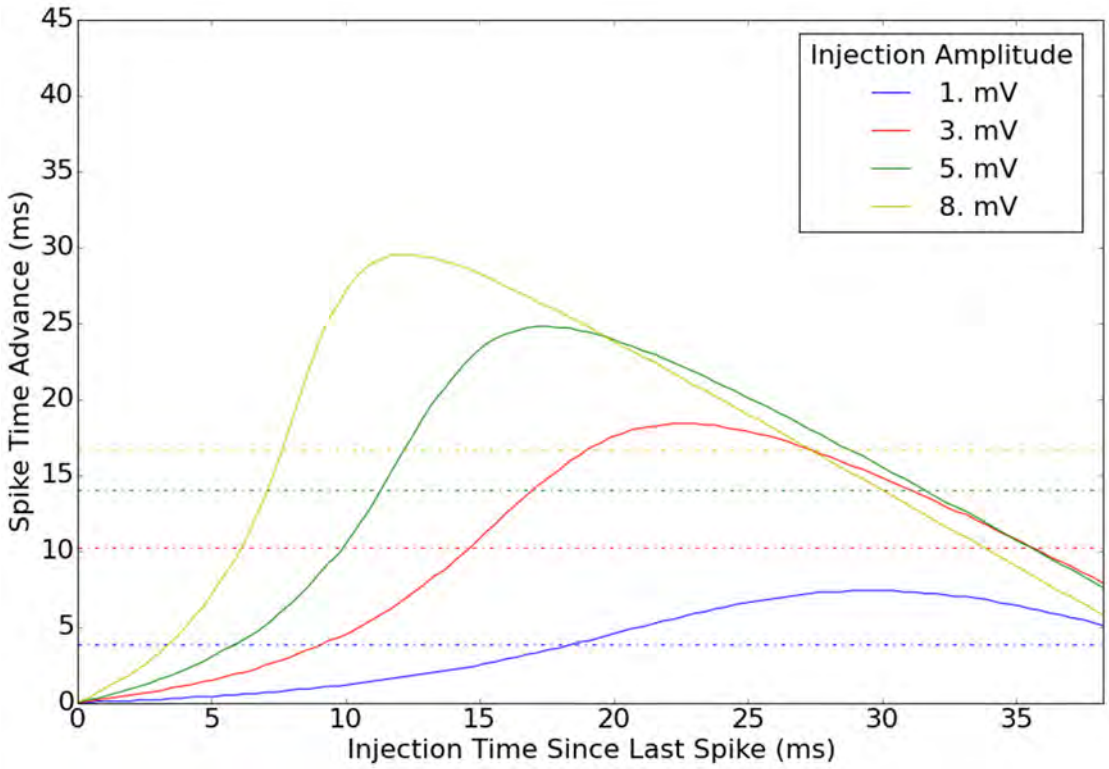


Figure 2-S 2 The phase/spike-time response curve. The curve showed the relationship between the spike time advance and current injection time since last spike. The average spike time advance (the dash lines) increases with more current and the cross point between the curve and average spike time advance is shifted toward the left side which suggested that faster conduction delay is required to achieve the spike-time based selectivity in the proposed computational principle.

Spike-timing dependent plasticity can enhance the spike-time based selectivity

Another well recognized computational principle in the neocortex is spike-timing dependent plasticity (Bi and Poo, 1998; Abbott and Nelson, 2000; Song et al., 2000; Caporale and Dan, 2008). With empirical evidence, the principle states that if a presynaptic neuron is often active just before spiking in the postsynaptic neuron, the synaptic weight between the two increases; on the other hand, if the presynaptic neuron is active just after the postsynaptic, the synaptic weight decreases. In the computational principle presented before, feedback always arrives just after a predictable neuron's action potential (since the feedback is caused by the predictable neurons) and will tend to arrive just before an unpredictable neuron's action potential (because the feedback itself tends to drive their activity). Thus, the feedback weight to the predictable neurons should decrease and feedback weight to the unpredictable neurons should increase (Figure 2-5 A). In such a situation, the proposed spike-time based selectivity should be enhanced.

We tested this idea by implementing STDP rules at the feedback synapses in the original 3-neuron model. The STDP followed the classical additive weight update rule (Song et al., 2000): the weight for the synapses increased and decreased in an exponential fashion. In the learning simulation, we used: the weight for pair (i, j) increased and decreased for the postsynaptic and presynaptic spike from i to j, respectively:

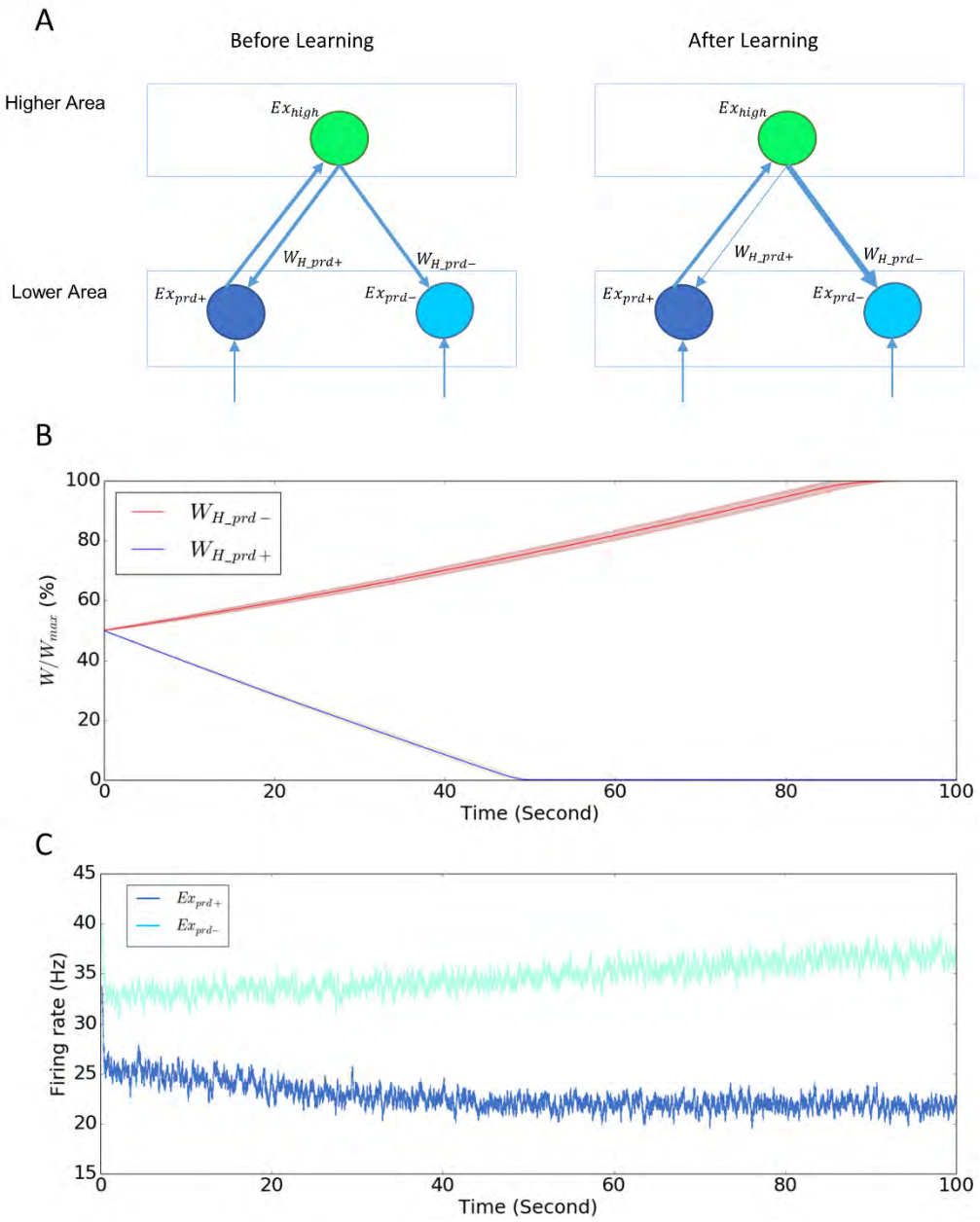


Figure 2-5 Spike-time based selectivity can be enhanced by the spike timing dependent plasticity. (A) The schematic diagram of the changing weight. In the initial state, the feedback weights are the same to both predictable neurons and unpredictable neurons ($W_{H\_prd+} = W_{H\_prd-}$). After learning, the feedback weights to predictable neurons increased and to unpredictable neurons decreased. (B) The relationship between the weight and time. The initial weights for W_{H_prd+} and W_{H_prd-} were set to the same value (50% of the

maximum weight). In the learning stage, the W_{H_prd+} decreased and W_{H_prd-} increased. The decreasing rate is higher than the increasing rate. (C) The firing rate of predictable neurons decreased and unpredictable neurons increased in the learning period.

Conclusion

In this chapter, we asked the question: how to create a better model of predictive coding based on the neurophysiological facts that most inter-areal feedback connections are excitatory and target excitatory neurons, and feedback usually exert a divergent connection pattern. We solved this question in a creative way by proposing a computational principle of spike time based selectivity: in a non-selective excitatory feedback network, the spike times of the higher area neuron are causally related to the spikes times of the predictable neurons in the lower areas, thus, the non-selective excitatory feedback will turn to a selective one due to the spike-time (different spike time advance in different positions in the phase response curve). The parameter setting simulations suggested that the proposed mechanism is biologically plausible and very robust. The balanced excitation and inhibition will turn the relative inhibition (relative to unpredictable neurons) into absolute inhibition (compare the situation with and without predictive feedback).

We also showed that if we apply the classical STDP rules to the neurons in this network, the non-selective excitatory feedback will turn to a selective excitatory feedback network where the excitatory feedback weight to predictable neurons will decrease and the excitatory feedback weight to

unpredictable neurons will increase. This STDP based dynamic will enhance the spike-time based selectivity.

This combination of the observed network structure and fundamental neuronal property made a convincing and probably universal computational principle. Indeed, we need more empirical evidence to prove this principle. However, if this principle is true, we will have a much deeper understanding of the working principle of the brain.

Chapter II

Predictive coding is an exciting field of research since there are many advantages in this theory: (1) predictive coding is based on the theory of efficient coding, which is accepted as the design principle of the brain by many researchers. (2) some of the empirical evidence supported the inhibitory feedback as described in the predictive coding theory. (3) many researchers describe predictive coding as "the model" of the brain and the future of the field of neuroscience.

As a promising model, predictive coding requires empirical evidence to support it. Indeed, there are some empirical evidence supporting the idea of predictive coding, especially about the inhibitory feedback. As I reviewed in the Introduction part of this thesis, the evidence on predictive coding mostly use the fMRI method. For example, Murray et al used three experiments to try to prove that shape perception could reduce activity in V1 (Murray et al., 2002). Further experimental work has linked predictive coding with expectation (Summerfield and Egner, 2009; Alink et al., 2010; Todorovic et al., 2011; Summerfield and de Lange, 2014), repetition suppression (Summerfield et al., 2008, 2011; Todorovic et al., 2011) and etc.

However, if we consider the predictive coding as a universal model for the interactions between the hierarchical areas, there are lots of problems with the original predictive coding model:

(1) In the neural circuits level, as described in the previous chapter, the original predictive coding does not fit the neurophysiological observation in the

excitatory role of feedback connections and the divergent connection patterns.

(2) At the neuronal populations level (psychophysics, fMRI, EEG, or MEG), one biggest problem is how to reconcile the predictive coding and attention since attention is usually considered as a feedback process and widely accepted as an excitatory role in the hierarchal brain. Empirical evidence usually shows an excitatory effect with attention.

Facing these problems about predictive coding, researchers usually use two types of strategies:

(1) Treating the predictive coding and the observed opposite evidence as fundamentally different mechanisms. For the difference between the original predictive coding model and observed neural circuits properties, researchers may argue that there is an intermediate stage (which usually is treated as a magical black box) between the neuronal population and neural circuits and the observed properties of neural circuits do not apply to the predictive coding which is supposed to be a population behavior. For the difference between attention and predictive coding, researchers may argue they use different neuronal populations or connections to realize them. Thus, for example, any observed excitatory feedback effects would be treated as evidence of attention, but inhibitory feedback effects would be treated as evidence of predictive coding.

(2) Try to reconcile predictive coding with only parts of observed evidence (which is supported by the original predictive coding model) and ignore other evidence (which is not supported by the original predictive coding model). For example, some researchers argued that predictive feedback

can be formed either about the content (leading to explaining away of incoming input, corresponding to weaker evoked responses) or the precision of lower-level input (leading to positive modulatory effects on the evoked responses, akin to attention). However, the fact that attention can also increase the lower-level activity (which is well supported by the biased competition theory) is ignored.

In this thesis, we used a very different strategy to face the problems in the original predictive coding theory: theory must follow the facts, not vice versa.

Thus, instead of trying to find out new evidence that supports predictive coding theory, we modified the original theory itself and proposed a model based on the idea of spike-time based selectivity: in a non-selective excitatory feedback network, the spike times of the higher area neuron are causally related to the spikes times of the predictable neurons in the lower areas, thus, the non-selective excitatory feedback will turn to a selective one due to the spike-time (different spike time advance in different positions in the phase response curve).

Under the proposed model, the feedback should not exert only one type of roles. However, in the predictive coding related evidence, we only see the inhibitory role of predictive feedback. Since these kinds of experiments only used the fMRI method which does not have a good temporal resolution, one possible reason for not detecting an excitatory effect of predictive feedback may be caused by the method. Thus, we used a psychophysical method to reinvestigate one of the first evidence about the inhibitory predictive feedback effect. Since the psychophysical method has a good temporal resolution, this study provides more information about predictive coding and its effect.

Shape perception enhances perceived contrast: evidence for excitatory predictive feedback?

Abstract

Predictive coding theory suggests that target-related responses are "explained away" (i.e., reduced) by feedback. Experimental evidence for feedback inhibition, however, is inconsistent: most neuroimaging studies show reduced activity by predictive feedback, while neurophysiology indicates that most inter-areal cortical feedback is excitatory and targets excitatory neurons. In this study, we asked subjects to judge the luminance of two gray disks containing stimulus outlines: one enabling predictive feedback (a 3D-shape) and one impeding it (random-lines). These outlines were comparable to those used in past neuroimaging studies. All 14 subjects consistently perceived the disk with a 3D-shape stimulus brighter; thus, predictive feedback enhanced perceived contrast. Since early visual cortex activity at the population level has been shown to have a monotonic relationship with subjective contrast perception, we speculate that the perceived contrast enhancement could reflect an increase in neuronal activity. In other words, predictive feedback may have had an excitatory influence on neuronal responses. Control experiments ruled out attention bias, local feature differences and response bias as alternate explanations.

Introduction

Predictive coding is a form of efficient sensory coding(Barlow, 1961b) that relies on the elimination of predictable neuronal responses and thereby the exclusive processing and transmission of unpredicted portions of the sensory input(Koch and Poggio, 1999; Rao and Ballard, 1999; Friston, 2005; Clark, 2013). As such, predictive coding could have important implications for the dynamics of information flow among the different levels of a sensory hierarchy such as the visual cortex.

Standard neuronal implementations of predictive coding assume that the feedback connections carry predictions of expected neural activity and the feedforward connections carry the residual activity between the predictions and initial lower area activity. To carry the residual, the feedforward connections are supposed to be excitatory, whereas to produce the residual the feedback connections are supposed to be inhibitory(Rao and Ballard, 1999; Friston, 2005). To simplify, standard neuronal models of predictive coding hold that the different hierarchical levels interact by excitatory feedforward carrying residual activity and inhibitory feedback carrying predictions. Recent implementations of predictive coding have divided neurons in each cortical area into two sub-populations, one coding for predictions/representations and one for prediction errors(Friston, 2005; Spratling, 2008a). These models suggested that only error units would be suppressed through either direct or indirect inhibition from the prediction units; the prediction/representation units, on the other hand, may actually be enhanced by predictive feedback(Spratling, 2008a, 2008b). Since theory must follow fact, it appears important to investigate the overall perceptual effect of feedback in predictive coding: is it excitatory or inhibitory?

Neurophysiology and neuroimaging provide converging supporting evidence for the hierarchical structure and excitatory feedforward connections of predictive coding models(Van Essen and Maunsell, 1983; Girard and Bullier, 1989; Girard et al., 1991b; Coogan and Burkhalter, 1993), but the experimental data are less unanimous regarding the inhibitory or excitatory nature of predictive feedback(Bastos et al., 2012): most neuroimaging studies show reduced activity by predictive feedback(Murray et al., 2002; Harrison et al., 2007; Summerfield et al., 2008; Alink et al., 2010), while neurophysiology indicates that most inter-areal cortical feedback is excitatory and targets mostly on the lower area excitatory neurons(Sandell and Schiller, 1982; Shao and Burkhalter, 1996; Johnson and Burkhalter, 1997; Hupé et al., 1998; Wang et al., 2000; Liu et al., 2013). In summary, the experimental literature does not clearly and unambiguously support the notion of inhibitory feedback, which is nonetheless an integral part of many models of predictive coding.

Here, we employed a psychophysical approach to investigate the properties of predictive coding. To produce predictive feedback, we employed similar stimuli as in Murray et al.: 3D-shape outlines and random-lines versions of the same stimuli(Murray et al., 2002). The former can be easily recognized, and should thus normally produce more predictive feedback than the latter. The two kinds of stimuli (3D shape and random lines) were displayed on gray disks simultaneously on the left and right of a fixation point on a black background. Subjects were asked to compare the luminance of the two disks (report the side of the brightest disk). The 3D-shape disk was perceived systematically brighter than the random-lines disk. Since there is experimental evidence suggesting a monotonic relationship between perceived contrast and neuronal activity in early visual areas(Dean, 1981; Boynton et al., 1999), we speculate that, at least at the moment at which subjects made their

perceptual decision about local contrast, predictive feedback was excitatory rather than inhibitory.

Results

Main Experiment: luminance judgment.

Participants (N=14) were instructed to fixate on the fixation point and judge the luminance of two gray disks on a black background on the left or right of fixation; each disk had either a 3D-shape or a random-lines pattern (randomly assigned) superimposed in its center (Figure 1, A). As these stimuli differentially activate higher visual areas (such as the lateral occipital complex, LOC(Murray et al., 2002)), one can reasonably expect different amounts of predictive feedback for the two locations(Murray et al., 2002), with more feedback towards the 3D-shape disk. Since anatomical evidence shows that feedback connections are strongly divergent(Salin and Bullier, 1995), we reasoned that the influence of predictive feedback might be measurable over the entire disk. We thus asked the participants to report the side of the disk that they perceived as brighter (after the stimuli offset, they received the instruction "which disk was brighter?", and responded via button press).

In each block of trials, one disk type was assigned with a fixed luminance value, while the other disk was assigned with a variable value around that level, different on each trial. The positions of the fixed-luminance and variable-luminance disks (and thus of the 3D-shape and random-lines stimuli) were randomly assigned in each trial. Two psychometric functions were computed from the data, one for blocks in which the random-lines disks had variable luminance values, and one for the other block type in which the 3D-shape disks

had variable luminance values. We finally compared these two psychometric functions: the psychometric shift was defined as the difference between the two psychometric thresholds (variable luminance value at which selection probability reaches 50%). A positive psychometric shift would suggest that the luminance of the random-lines disk at which it is perceived equiluminant to the fixed-luminance 3D-shape disk is higher than the luminance of the 3D-shape disk at which it is perceived equiluminant to the fixed-luminance random-lines disk. In simpler terms, a positive effect indicates that 3D-shape disks are perceived as brighter than random-lines disks, while a negative effect implies the opposite relation.

Results showed a positive effect for all 14 subjects, i.e. they perceived 3D-shape disks brighter than random-lines disks (Figure 1, B-C). The psychometric shift was 8.04% ± 2.82% (average ± standard deviation across subjects) normalized luminance and the grand average psychometric shift (when pooling data over all subjects) was 7.93%. A student's paired t test for the psychometric shift shows $t(13)=10.69$, $p < 8.29\times10^{-8}$ with a confidence interval of (6.42%, 9.67%). This effect was unlikely to be due to eye movements or faulty fixation: in two subjects (indicated in Figure 1.C by colored bars) eye position was monitored by an eye-tracker and any trial with sizeable eye movements were discarded; these two subjects still produced positive psychometric shift (3.57% and 5.51%) that were well within the range of the group. Since luminance/contrast discrimination judgments are linked to neuronal activity in early visual cortical areas(Dean, 1981; Boynton et al., 1999), these results indicate that at the moment at which subjects made a decision about luminance/contrast discrimination, the 3D-shape had presumably produced more neuronal activity in early cortical areas than the random-lines stimulus. As the 3D-shape is more recognizable than the random-lines and thus more likely to induce predictive feedback signals, we tentatively conclude that predictive

feedback had an excitatory effect on neuronal activity in early visual cortex. However, we also tested several alternative explanations.

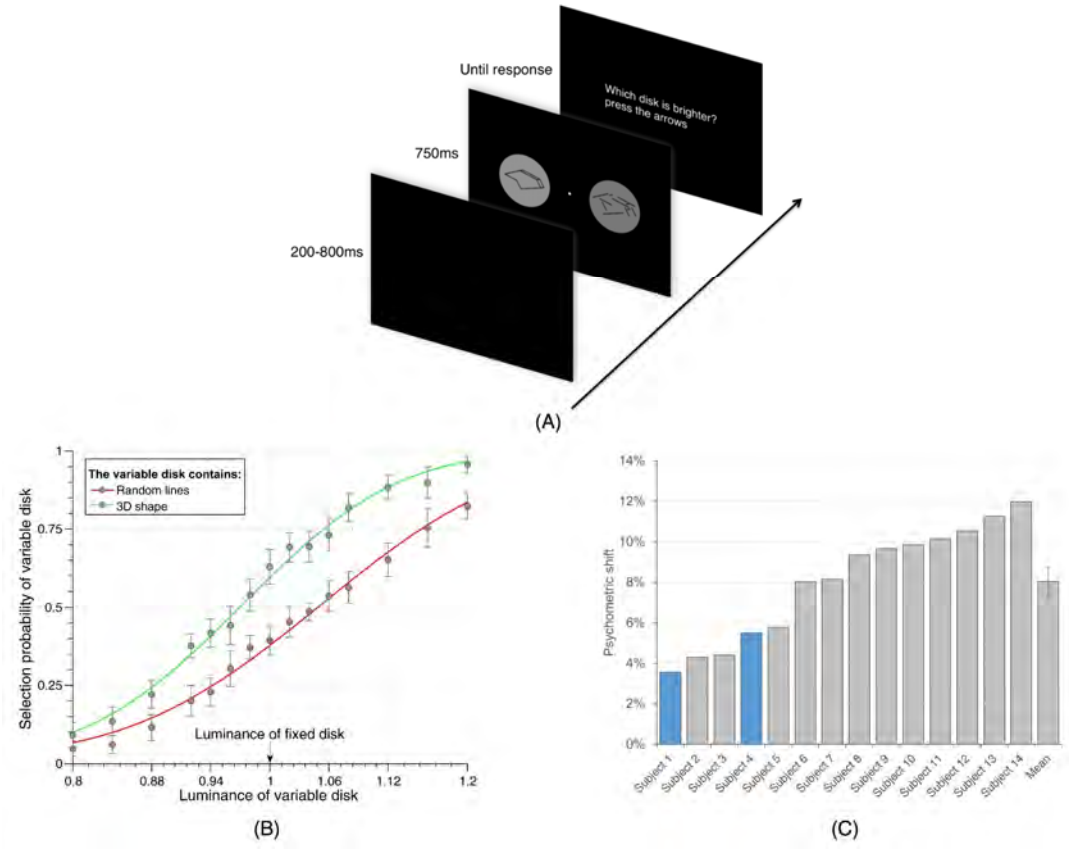


Figure 3-1. Main experiment and results. (A) Experimental paradigm. Each trial consisted of a 200-800ms blank screen, a 750ms stimulus screen and a response screen that remained visible until the response was provided. The stimulus screen consisted of a fixation point, one circular gray disk with a 3D-shape stimulus and another with a random-lines stimulus (with randomized positions on the left and right of fixation for every trial). One disk had a fixed contrast level and the other a variable contrast value around that level (randomly assigned on every trial). Subjects were instructed to compare the luminance of the two disks. No feedback was given after the response. (B) Comparison of the grand average psychometric functions (when pooling data over all subjects). Each curve represents the selection probability of the variable disk when this disk contained the 3D-shape (green) or the random-lines stimulus (red). Error

bars represent standard error of the mean (SEM) across subjects (C) Psychometric shift for each subject and mean across subjects. Psychometric shift was defined as the difference between the two psychometric functions at 50% selection probability. All 14 subjects showed a positive effect, with the disk behind the 3D-shape stimulus perceived brighter against the black background than the one behind the random-lines. Subjects 1 and 4, marked by colored bars, performed the experiment while their eye position was monitored, and any eye movement or break of fixation discarded. Error bar represents SEM.

Control experiment: attention bias.

An obvious possible confound with our experimental design could be a systematic attention bias towards 3D-shape stimuli. Indeed, previous fMRI studies showed that attention can increase activity in early visual cortical areas(Corbetta et al., 1995; Gandhi et al., 1999) and alter stimulus appearance including perceived contrast(Carrasco et al., 2004). Is the enhanced perceived contrast for 3D shapes simply a product of increased attention? If this was the case, then one would expect the psychometric shift to decrease when attention is diverted from the peripheral disks using a challenging central task(Corbetta et al., 1991). We thus replaced the fixation point with a rapid serial visual presentation (RSVP) stream of letters. The observers (a subset of participants from the main experiment; N=7) were instructed to count the number of occurrences (from 1 to 4) of the letter "T" (Figure 2, A), a task known to demand important attentional resources(Joseph et al., 1997; Braun, 1998). To ensure that attention was properly engaged by this central task, we used a presentation speed (6.67 letters/s) that made the task highly challenging (correct rate, 72.48% ± 12.37%, average ± standard deviation across subjects). Participants were instructed to prioritize the counting task and to respond to it

first; negative auditory feedback was given after every mistake in this counting task.

Two psychometric functions were generated using the same method as in the main experiment, and compared with the psychometric functions obtained from the same participants during the main experiment (Figure 2, B-C). The psychometric functions had significantly shallower slopes (as measured by the standard deviation of a fitted cumulative normal distribution) than in the main experiment (attention bias control *vs.* main experiment 0.33 ± 0.13 *vs.* 0.16 ± 0.07, average ± standard deviation, $t(13)=4.7$, $p<4.23\times10^{-4}$, the psychometric function with 3D-shape disk as the variable-luminance disk and the psychometric function with random-lines disk as the variable-luminance disk were analyzed jointly, and 14 pairs of standard deviation values were thus compared for the analysis), suggesting that attention was properly engaged and that subjects were therefore less sensitive to contrast differences(Corbetta et al., 1991). Given that attention was significantly engaged in the central counting task, and regardless of the magnitude of this engagement (i.e., even if only a portion of attentional resources was engaged), an attentional account of our previously observed contrast perception shift should predict that the shift would decrease during the dual-task condition. However, the psychometric shift was not decreased (if anything, it even increased marginally): across subjects, the psychometric shift for this control experiment was 9.27% ± 6.13% (average ± standard deviation across subjects) when including all trials, and 9.34% ± 6.86% when including only those trials in which the counting task was performed correctly (and thus attention was presumably more efficiently engaged); this is to be compared with a psychometric shift of 8.3% ± 3.2% during the main experiment. Paired t-tests showed that the result differences between the control experiment and the main experiment were not significant (including all trials *vs.* main experiment: $t(6)=0.467$, $p>0.65$;

counting task correct trials *vs.* main experiment: t(6)=0.407, p>0.69). The grand average psychometric shift (when the psychometric functions were computed from the grand-average data across the seven participants) was 8.82% for all trials, and 8.70% for counting-task correct trials, relative to a psychometric shift of 8.18% during the main experiment.

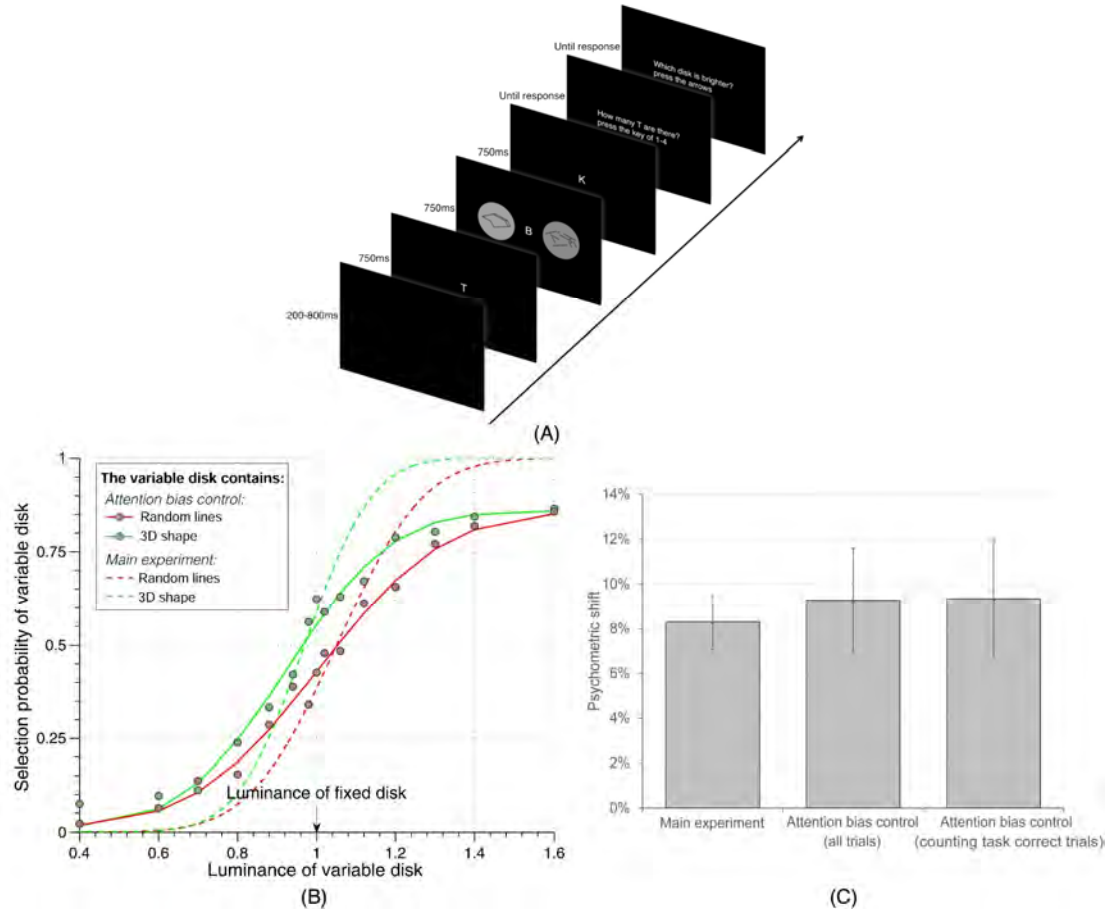


Figure 3-2 Attention bias control (A) Experimental paradigm. Each trial consisted of a 200-800ms blank screen, a 2250ms letter RSVP sequence in the center, a 750ms stimulus screen starting 750ms after the beginning of the letter RSVP, and two successive response screens, each presented until a response was provided. The RSVP sequence displayed randomly drawn letters every 150ms (the same letter could not appear twice in a row). The stimulus screen was the same as in the main experiment, expect for the replacement of the fixation point by the RSVP sequence. Subjects were

instructed to first count the number of letters "T" in the RSVP, and (as a secondary task) to compare the luminance of the two disks. Negative auditory feedback was given after every mistake in the counting task. (B) Comparison of grand average psychometric functions for the same subjects in the attention bias control (solid lines) and in the main experiment (dashed lines). (C) Comparison of psychometric shift for the same subjects in the main experiment and during the attention bias control, either including all trials, or only those in which the counting task was performed correctly. Similar psychometric shift was obtained for all conditions, indicating that attention bias is unlikely to explain our findings. Error bars represent SEM.

Control experiment: local features.

We tested yet another alternative interpretation: that low-level local features altered the perceived contrast. Even though the paired 3D-shape and random-lines stimuli have the same number of line segments, comparable line orientations, retinotopic distribution and overall luminance, they also differ in some respects, for example the presence of corners and line junctions in 3D-shapes only. It is conceivable that such local features could influence the processing of local contrast, and that in turn this local alteration of perceived contrast could propagate to the entire disk via filling-in mechanisms. This local contrast alteration mechanism, however, is different from the postulated excitatory feedback effect, since the latter is assumed to depend on the entire shape and thus to be more global in nature. Thus, the two alternative accounts make different predictions about the consequence of changing the contrast polarity of the stimulus outline (black *vs.* white) while keeping the disk luminance (gray) and the screen background luminance (black) unchanged. Indeed, if local features are affecting contrast perception locally, then a white outline on a gray disk (instead of a black outline on a gray disk, as in the main experiment) should result in a reversed contrast effect (3D-shape disk perceived darker than the random-lines disk). On the other hand, the effect of

global feedback should not solely depend on the luminance of the stimulus outline (black or white), but also on the contrast between the (gray) disk and its (black) background; if that contrast does not change, the effect of global feedback might be expected to decrease, but should not fully reverse. To distinguish between these alternatives, in this control experiment we replaced the black outline of the 3D-shape and random-lines with white outlines (keeping the disks gray and the screen background black), and asked subjects to perform the same comparison task as in the main experiment (judge which of the two disks is brighter).

We found that the effect was not reversed by the change of contrast polarity (Figure 3). The psychometric shift for this control experiment was 3.23% ± 4.44% (average ± standard deviation across subjects; N=10 including 4 participants from the main experiment); the grand average psychometric shift (when the psychometric functions were computed from the grand-average data across participants) was 3.1%. A one-sample Student's t-test showed that this effect was incompatible with a full reversal (null hypothesis of an psychometric shift of -8.04%, based on the results reported in Figure 1; $t(9)= 8.03$, $p < 2.15\times10^{-5}$); in fact, this effect was still greater than zero ($p < 0.05$). This implies that the local contrast polarity is not the sole determinant of the observed effect, and that global feedback must also contribute to it. Therefore, our interpretation of an excitatory feedback still remains viable.

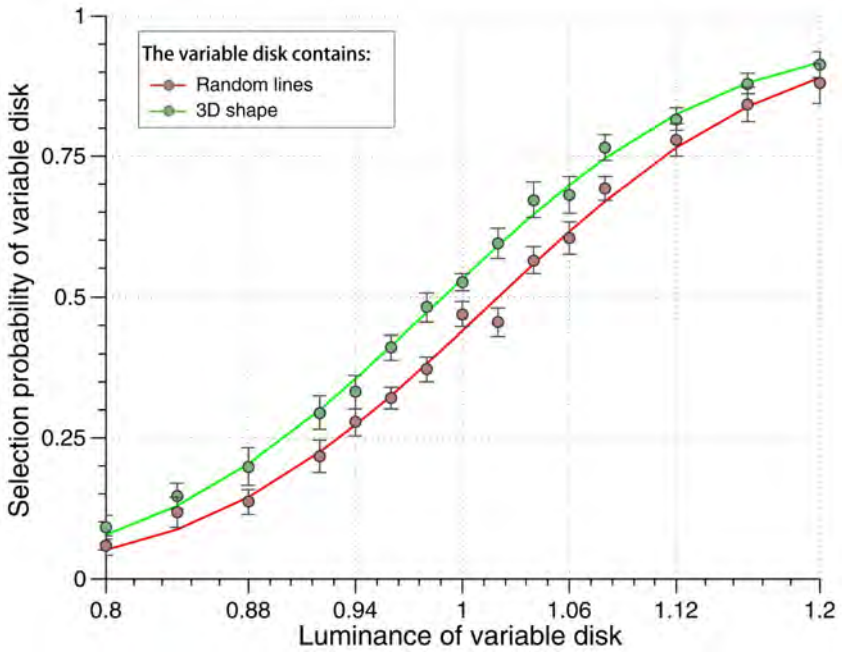


Figure 3-3 Comparison of the grand average psychometric functions in the "local features" control experiment. In this experiment, the contrast polarity of the stimulus outline was reversed (from black to white) to evaluate the contribution of local features on psychometric shift. While the grand average psychometric shift was reduced, it remained positive ($p<0.05$), and did not fully reverse ($p < 2.15\times10\text{-}5$) as would have been predicted if local features were responsible for the entire effect. Error bars represent SEM.

Control experiment: response bias.

We also tested the possible influence of a response bias. One might imagine that when observers do not truly perceive any contrast difference between the 3D-shape and the random-lines disks, but are still confronted with a forced choice between two responses, they could be inclined to systematically choose the one stimulus that they recognized (i.e. the 3D-shape). If this was the case, however, reversing the task instructions (asking "which disk was darker?" instead of "which disk was brighter?") should not affect this response bias, and

should thus produce a reversed psychometric shift (3D-shape disk perceived *darker* than random-lines disk). We re-tested seven participants from the main experiment using these reversed instructions (Figure 4). None of them showed a reversed effect. The psychometric shift was 8.11% ± 3.54% (average ± standard deviation across subjects), compared with 8.59% ± 2.75% for the same subjects during the main experiment. A paired t-test showed that the differences were not significant ($t(6)=0.2891$, $p > 0.78$). The grand average psychometric shift was 8.06% compared with 8.46 % for the main experiment. Thus, response bias is unlikely to account for our findings.

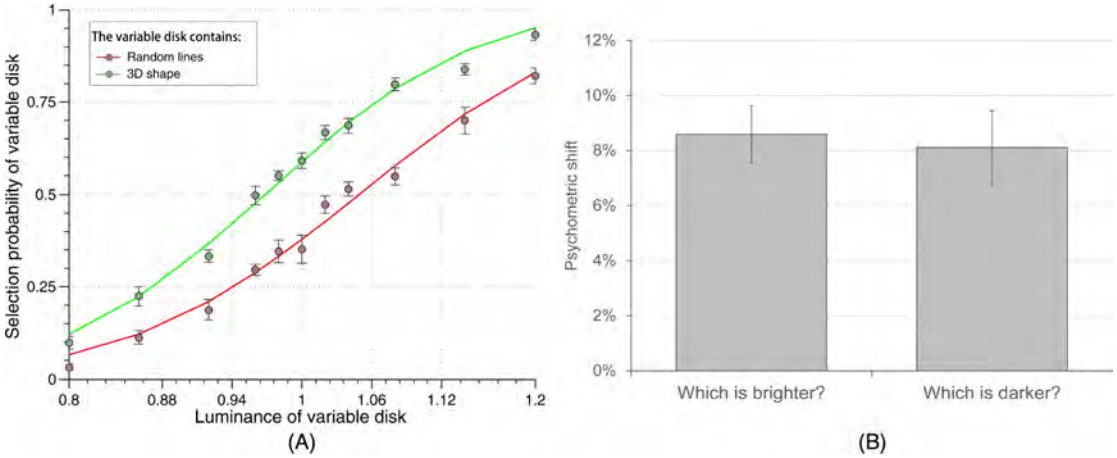


Figure 3-4 (A) Comparison of the grand average psychometric functions in the "response bias" control. (B) Comparison of mean psychometric shift for the same subjects in the main experiment and the "response bias" control. In this experiment, the response instruction was reversed (report the darker disk) to measure the influence of a possible response bias. Psychometric shifts were similar in the two conditions (t-test, $p > 0.78$), indicating that response bias is unlikely to play any major role in the effect. Error bars represent SEM.

Control experiment: same/different judgment.

Finally, as an even more stringent test against response bias, we instructed subjects (N=5) to perform a same/different luminance judgment task (asking "Did the two disks have the same luminance?" at the end of each trial). Any response bias towards either the 3D shape or the random lines stimulus would not be expected to affect responses in this sort of task. For different types of trials (3D-shape or random-lines inside of the variable-luminance disk), we measured the probability of "same luminance" response as a function of the luminance of the variable-luminance disk. If shape perception truly has an effect on contrast/luminance perception, we should expect a shift of the distributions. Indeed, we found a right-shift of the distribution of "same" responses when random lines were inside of the variable-luminance disk (relative to the distribution of "same" responses when 3D shape were inside of the variable disk), indicating that 3D shape enhanced perceived contrast/luminance (Figure 5). By fitting each distribution to a Gaussian function and comparing their peaks, we found an average psychometric shift of 5.10% ± 2.43% (average ± standard deviation across subjects). This psychometric shift corresponded to a p value of 0.0093 with a confidence interval of (2.08%, 8.12%). To compute the grand average psychometric shift, we first normalized the response distributions of each subject relative to their mean value across all possible variable luminance, and then we fitted the average normalized distributions with Gaussian functions. The grand average psychometric shift over 5 subjects was 3.98%. Since this measurement is less prone to response biases, we thus re-confirmed our findings with convergent evidence.

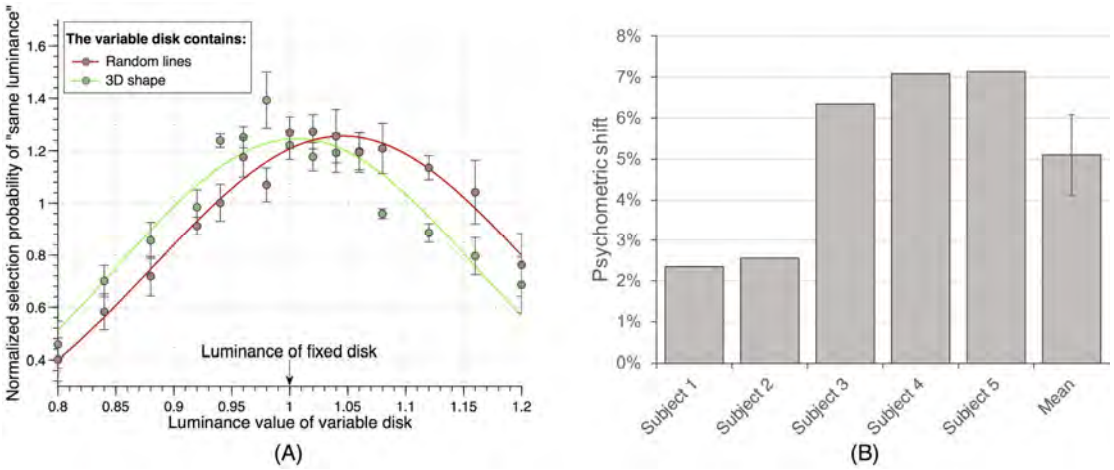


Figure 3-5 (A) Comparison of the grand average psychometric functions in the "same/different" experiment. In this experiment, we instructed subjects to report whether the two disks had the same or different luminance. By comparing the distribution of normalized "same luminance" responses (normalized by mean response probability) on different types of trials (3D-shape or random-lines inside of the variable-luminance disk), we could determine which disk was perceived brighter. The right-shift of the "same" response distribution with random lines inside of the variable disk (or the left-shift of the "same" response distribution with 3D shape inside of the variable disk) indicates that 3D shape enhanced perceived contrast/luminance. (B) Psychometric shift for each subject and mean across subjects. Psychometric shift was defined as the difference between the peaks of the two psychometric functions. All 5 subjects showed a positive effect, with a right-shift of the "same" response distribution with random lines inside of the variable disk. Error bar represents SEM across subjects.

Discussion

In the present study, consistent behavioral responses of 14 subjects (Figure 3-1) revealed that the disk behind the 3D-shape stimulus (which could be easily recognized, and give rise to predictive feedback) was perceived brighter against the black background than the one behind the random-lines (meaningless) stimulus. Given previous evidence suggesting a monotonic relationship between contrast perception and neural activity in early visual areas(Dean, 1981; Boynton et al., 1999), we tentatively interpret these results as evidence that predictive feedback had an excitatory effect on sensory activity, at least at the time point at which contrast perception was established.

We performed four control experiments to rule out alternative explanations of our results. By replacing the center fixation point with an attentional demanding task (letter RSVP), we obtained similar psychometric shifts for all conditions, indicating that attention bias was unlikely to explain our findings (Figure 3-2). In the main experiment, two contrasts could have contributed to the perceived disk luminance: a local one reflecting the luminance difference between stimulus lines and disk, and a more global one caused by the luminance difference between disk and screen background. Both contrasts could have been affected by predictive feedback (e.g., due to divergent feedback connections); but in addition, the local contrast could also have been modulated by more local confounding factors, such as systematic physical differences in the random lines *vs.* 3D-shapes stimuli (although the number of lines and corresponding numbers of pixels were equated, higher-order statistics reflecting inter-pixel relations were not equated). To test if the local factors could solely account for our results, we examined the relative contribution of local and global contrast to the perceived disk luminance by reversing the polarity of the stimuli outline, from black to white. This operation

reversed the direction of the contribution from local contrast: if it had previously resulted in the disk being perceived brighter, it should have now caused it to be perceived darker. We showed, however, that psychometric shifts did not fully reverse, indicating that local factors were unlikely to explain all of our findings (Figure 3-3). Finally, we used two separate experimental manipulations to assess the effect of response biases on our results: we modified the response instructions (asking "which disk was darker?" instead of "which disk was brighter?", Figure 3-4), and in a separate control we changed the subjects' task (to a same/different perception task, by asking "Did the two disks have the same luminance?", Figure 3-5). The comparable psychometric shifts obtained regardless of task instructions indicated that response biases were unlikely to explain our findings.

These results concur with neurophysiological evidence that cortico-cortical feedback connections are mainly excitatory(Sandell and Schiller, 1982; Shao and Burkhalter, 1996; Johnson and Burkhalter, 1997; Hupé et al., 1998; Wang et al., 2000). However, they also appear to contradict neuro-imaging evidence suggesting that predictive feedback is inhibitory, using a similar paradigm and the same set of stimuli as in the present study(Murray et al., 2002). The major difference between our study and that of Murray et al.(Murray et al., 2002) is the dependent variable used to estimate neural activity: perceived contrast *vs.* BOLD activity. The existence of a monotonic relationship between contrast and neural activity in early visual cortical areas has been well established in neurophysiology(Dean, 1981). The contrast response function of striate cortex neurons has been directly measured in cat and monkey(Albrecht and Hamilton, 1982). In human primary visual cortex, contrast is directly related to BOLD responses(Goodyear and Menon, 1998), and psychophysical contrast judgments (i.e., perceived contrast) are also linked to BOLD responses in visual areas V1, V2d, V3d and V3A(Boynton et al., 1999). Selective contrast tuning

exists for some V4 neurons, however, contrast still has a monotonic and positive relationship with the activity of overall V4 neuron populations(Sani et al., 2013). It thus appears reasonable to use perceived contrast as a proxy for overall neuronal activity in early visual cortex. On the other hand, perceived contrast and BOLD activity certainly differ in terms of their temporal resolution: perceptual decisions can be made within a few hundred milliseconds, whereas BOLD signals have a slower time course and a much poorer temporal resolution (on the order of seconds) due to the nature of the hemodynamic response function. Thus, it is possible to envision that predictive feedback could play an excitatory role during early stages of stimulus processing, and yet have a long-lasting inhibitory effect on subsequent neuronal activity.

With the same set of stimuli but complementary methods, the combination of our psychophysical study and previous neuro-imaging results(Murray et al., 2002) thus highlights a possibly more comprehensive temporal profile for predictive feedback. But, is this profile universal? Is it comparable across all brain regions? Summerfield et al. and Egner et al. investigated predictive feedback by measuring BOLD responses in FFA(Summerfield et al., 2006; Egner et al., 2010). With 750ms-long face images, Egner et al. showed that FFA responses decreased with high prior expectation compared to low expectation. On the other hand, with masked 100ms-long face images, Summerfield et al. found that FFA responses increased during a face-related task compared to a non-face-related task. Even though none of these authors explicitly linked these two studies with respect to stimulus timing, the corresponding time-line of predictive feedback in FFA appears compatible with our hypothesis. At the opposite end of the visual system, Olsen et al. showed that the corticothalamic feedback from layer 6 of mouse V1 to lateral geniculate nucleus (LGN) played an inhibitory role: a large proportion of visually evoked activity in LGN relay neurons was inhibited when driving V1

layer 6 neurons optogenetically(Olsen et al., 2012). Nonetheless, anatomical evidence suggests that direct feedback connections from visual cortex to LGN relay cells are actually excitatory(Guillery and Sherman, 2001), but visual cortex also sends excitatory feedback to the thalamic reticular nucleus (TRN), a layer of inhibitory neurons adjacent to the thalamus, which can in turn inhibit LGN relay neurons. It thus seems plausible that direct corticothalamic excitatory feedback might influence LGN relay cells before the arrival of indirect inhibitory feedback from the TRN. Thus, even for connections between other areas than V1 and extrastriate visual cortex, predictive coding may present the same hypothesized temporal profile: excitation followed by inhibition.

Furthermore, even though inter-areal feedback connections are carried out only via the excitatory neurons (since only they have long enough axons to connect different areas) and mostly target excitatory neurons(Johnson and Burkhalter, 1996, 1997), the net effects of feedback are not always excitatory(Bastos et al., 2012). Hupé et al. showed that with very low saliency stimuli, cooling down V5, and thus interrupting its feedback, actually increased neural activities in V3(Hupé et al., 1998). Schneider et al. also revealed inhibitory effects of feedback in auditory cortex(Schneider et al., 2014). One possible mechanism for such inhibitory effects is excitatory cortico-cortical feedback reducing lower level activities by activating local inhibitory circuits(Schneider et al., 2014; Zhang et al., 2014). This possible mechanism may help us reconcile our findings with neuroimaging results: one group of neurons in early visual cortex may be excited by the top-down prediction (i.e. the 3D shape); this enhancement could in turn activate the local inhibitory circuits to inhibit other groups of neurons, leading to an overall inhibitory effect. Since the excited neurons and the inhibited ones belong to different populations, this mechanism might result in a spatial dissociation of excitatory and inhibitory effects (rather than, or in addition to, the postulated temporal dissociation).

Kok et al. provided evidence for such a spatial dissociation: they observed enhanced activity in the area where a Kanisza-like illusory shape was perceived, but reduced activity for the surrounding inducers(Kok and de Lange, 2014).

Predictive coding is a powerful scheme that describes perception as an inferential process "explaining away" predicted responses from input signals(Rao and Ballard, 1999; Friston, 2005). However, only limited experimental observations on this phenomenon are available. Based on these limited observations, several neuronal models of predictive coding have been put forward. Friston et al. (Friston, 2005; Friston and Kiebel, 2009) improved on Rao and Ballard's original predictive coding model(Rao and Ballard, 1999) and proposed a specific distribution of functional roles across the cortical layers(Mumford and Mumford, 1992). Spratling(Spratling, 2008a) advocated a neuronal model with excitatory feedback which, according to our logic described before, fits better with the anatomical and neurophysiological evidence(Sandell and Schiller, 1982; Shao and Burkhalter, 1996; Johnson and Burkhalter, 1997; Hupé et al., 1998; Wang et al., 2000).

As pointed out already by Spratling(Spratling, 2008a), one possible way to dissolve the conceptual tension between classical models of feedback (e.g. biased competition) and predictive coding is by hypothesizing that all predictive coding schemes employ two types of neurons within each layer of the cortical hierarchy: prediction or representation units (P) and prediction error units (E). Feedback aims to inhibit the error units, but thereby also strengthens the representation at the lower level. Under some simplifying assumptions, this hypothesis makes classical models of biased competition and predictive coding mathematically equivalent(Spratling, 2008a). In line with this notion, Kok et al. observed reduced overall activity for expected stimuli, yet an

increased stimulus representation(Kok et al., 2012b). These findings are inconsistent with the idea that feedback globally inhibits sensory representations; rather, they support the notion that it is only the error units that are suppressed, and thereby predictions increase the signal-to-noise ratio.

In conclusion, the present psychophysical study showed an excitatory influence of predictive feedback at the perceptual level. To build an optimal neuronal model of predictive coding, the consideration of the entire range of neuroimaging, neurophysiological and psychophysical evidence is necessary. We hope the observed excitatory influence of predictive feedback could thus help improve the design of future predictive coding models.

Methods

Subjects

Based on pilot experiments, we expected an average psychometric function shift of at least 5%, with a variability of 5% in the point of subjective equality of psychometric functions. To reach a statistical power of at least 95%, we determined that the sample size was twelve subjects. To monitor eye movements, we added two more subjects with an eye-tracker. Finally, fourteen volunteers (7 female, mean age 27.78 ± 3.78 years, one left handed, five with left eye dominance) participated in the main experiment.

Seven of these main experiment participants (4 female; mean age 28.5 ± 1.2 years; all right handed) performed the attention control experiment. Four main experiment participants and six other volunteers (10 participants, 5 female, mean age 28.1 ± 4.9years) performed the “local features” control experiment. Seven main experiment participants (3 female, mean age 29.6 ± 4.2 years) performed the “response bias” control experiment. Two main experiment participants and three other volunteers (5 participants, 3 female, mean age 28.8 ± 2.2 years) performed the “same/different” control experiment. These sample sizes for control experiments were determined, based on the effect size obtained in the results of the main experiment, so as to ensure a minimum statistical power of 80% for each control experiment.

All subjects in the main experiment and all control experiments had normal or corrected to normal vision. The study was approved by the local ethics committee “Sud-Ouest et Outre-Mer I” and followed the Code of Ethics of the

World Medical Association (Declaration of Helsinki). All subjects provided signed informed consent before starting the experiments.

Apparatus

Stimuli were presented at 57 cm distance using a desktop computer (2.09 GHz Intel processor, Windows XP) with a cathode ray monitor (resolution: 800×600 pixels; refresh rate: 120 Hz, Gamma corrected luminance function). Stimuli were designed and presented via the Psychophysics Toolbox(Brainard, 1997) running in MATLAB (MathWorks).

Stimuli and tasks

Twenty pairs of 3D-shape and random-lines stimuli were first generated. Similar to Murray et al. (2002), 3D-shapes were generated by randomly selecting 4–6 vertices, connecting the vertices and adding small extensions to render perceived depth. Random-lines stimuli were created by breaking the 3D-shape at its intersections and randomly shifting the lines (crossings were avoided) within the display. The diameter of both 3D-shape and random-lines stimuli was 3 degrees. In all experiments except the “local features” control experiment, the stimulus outlines were black. In the “local features” control experiment, these outlines were white.

Main experiment. Stimuli consisted of a central white fixation point (diameter: 0.2 degrees of visual angle) and two circular gray disks (diameter: 4 degrees each). One 3D-shape stimulus was in the center of one disk (3D-shape disk) and one random-lines in the other (random-lines disk). The disks were presented at an eccentricity of 3 degrees randomly on either side (left or right)

of fixation. The luminance of the disks ranged from 20.17 cd/m^2 to 32.3 cd/m^2 (measured with a Minolta Chroma Meter CS-100, Minolta Co., Ltd, Osaka, Japan). To compute normalized luminance, the measured luminance values were divided by the middle value of the luminance range, i.e. 26.235 cd/m^2. One disk had a fixed luminance level (100% normalized luminance) and the other a variable luminance value, randomly drawn from the normalized luminance set [80%, 84%, 88%, 92%, 94%, 96%, 98%, 100%, 102%, 104%, 106%, 108%, 112%, 116%, 120%]. Disks were presented on a black background (normalized luminance 0.0145 %). Before stimulus onset, there was a blank screen that lasted from 200 to 800ms (random uniform distribution). The stimulus lasted for 750 ms and then the instruction "Which disk was brighter? Press the arrows" appeared on the screen until the subject's response. Subjects were presented with 5 blocks of 200 trials, and asked to fixate the fixation point and use the arrow keys on a standard 105 key keyboard to respond (left arrow for left is brighter, right arrow for right is brighter). There was no feedback after the response. To monitor for breaks in fixation, eye movements of two subjects were recorded using a video-based eye tracker (EyeLink 1000 plus, SR Research, Ontario, Canada) with a sampling rate of 1000 Hz. The eye tracker was calibrated at the beginning of each block (only 4 blocks of 200 trials were performed by these subjects). For each trial, if the maximal deviation from fixation during stimulus presentation was bigger than 0.5 degrees from the fixation point, the trial was rejected automatically and the instruction "Please fixate on the fixation point" appeared on the screen.

Control experiment: attention. The fixation point was replaced by a rapid serial visual presentation (RSVP) stream of letters. The RSVP was made up of letters randomly drawn from the set [T, L, K, J, B, C, D], 2 degrees in diameter. Each letter was presented for 150 ms. A letter could not appear twice in a row, and the letter "T" appeared from one to four times (randomized from trial to trial).

The RSVP sequence started before the disks presentation and ended after the disks. Fifteen letters were presented from time 0 to 2250 ms, while the disks were presented from time 750 to 1500 ms. The instruction "How many Ts were there? press the key of 1-4" appeared at time 2250ms, until the subject's response (using the keys 1,2,3,4 in the numeric keypad). There was a short beep sound feedback if subjects answered incorrectly in this task. After the subject's response to the letter counting task, the instruction "Which disk was brighter? Press the arrows" appeared, and subjects performed the luminance judgment task as in the main experiment. Subjects were presented with 8 blocks of 100 stimuli. They were instructed that their primary task was to count the number of occurrences of the letter "T".

Control experiment: local features. Stimuli were white 3D-shape and random-lines on a gray disk. The task was the same as in the main experiment.

Control experiment: response bias. Stimuli were the same as in the main experiment. The variable luminance value was randomly drawn from the normalized luminance set [80%, 86%, 92%, 96%, 98%, 100%, 102%, 104%, 108%, 114%, 120%]. The question given after each trial was: "Which disk was darker? Press the arrows", and subjects were instructed to choose the darker disk using the arrow keys.

Control experiment: same/different judgment. Stimuli were the same as in the main experiment. The question given after each trial was: "Did the two disks have the same luminance?", and subjects pressed one key to indicate that they perceived the same luminance and another key when they perceived a different luminance.

Data analysis

The trials were classified into two categories: either the 3D-shape disk had a variable luminance value, or the random-lines disk had a variable luminance value. For all experiments except the same/different luminance experiment, for each trial type, the selection probability of the disk with the variable luminance value was computed, separately for each variable luminance value. Two psychometric functions were generated, one for each trial type, expressing the selection probability as a function of the variable luminance value, and fitted using normal cumulative distribution functions (each pair of psychometric functions was fitted with Gaussian cumulative functions with six parameters: mean and standard deviation separately for each psychometric function, and a common guess rate and lapse rate for both functions; the guess rate was set in the range of [0,1] and the lapse rate was set in the range of [0, 1-guess rate], which limited the maximum and minimum values of the psychometric functions to 1 and 0, respectively). Finally, we compared these two psychometric functions. The difference between the two psychometric functions at 50% selection probability was defined as the psychometric shift (and the difference between the two grand-average psychometric functions was defined as the average psychometric shift). A student's t test against the null hypothesis of a psychometric shift equal to zero (both disks perceived equally bright) was performed using psychometric shift from all subjects. For the same/different judgment, the probability of reporting “same luminance” was computed for each variable luminance value, and two psychometric functions were generated and fitted using Gaussian distribution functions, separately for each trial type. The difference between the peaks of the two Gaussian functions was defined as the psychometric shift in this experiment.

Conclusion

In this chapter, we reinvestigated one of the first evidence about the inhibitory predictive feedback effect: shape perception. By using the psychophysical method, we can obtain a much better temporal resolution than the traditional fMRI method. We obtained surprising and consistent results in the view of traditional predictive coding: we found out that shape perception can enhance the perceived contrast. Then we used the perceived contrast as a proxy for overall neuronal activity in early visual cortex and concluded that shape perception can increase the neuronal activity in early visual cortex. We performed control experiments to exclude three possible alternative explanations of our results: attention bias, local factors and response bias.

Our psychophysical study showed an excitatory influence of predictive feedback at the perceptual level. This result seems to be contradictory to the first inhibitory evidence of predictive feedback and suggested a different effect in perception and fMRI results. Since we used the same stimuli as the first inhibitory evidence of predictive coding and our control experiments on traditional attention effect, the evidence is hard to be treated as "attention" effect or ignored. I think the contradiction may reveal a rich profile of the predictive feedback and two potential possibilities may explain the observed contradictions:

(1) A comprehensive temporal profile for predictive feedback. Since perceived contrast and BOLD activity in fMRI differ in terms of their temporal resolution (a few hundred milliseconds for perception and several seconds for BOLD signal), it is possible to envision that predictive feedback could play an excitatory role during early stages of stimulus processing, and yet have a long-lasting inhibitory effect on subsequent neuronal activity.

(2) A comprehensive spatial profile for predictive feedback. Since neurophysiological evidence showed that the local inhibitory circuits can be activated by the excitatory cortico-cortical feedback and result in absolute inhibitory effect, it is possible that one group of neurons in early visual cortex may be excited by the top-down prediction and the activation of the local inhibitory circuits can inhibit other groups of neurons, leading to an overall inhibitory effect and result in a spatial dissociation of excitatory and inhibitory effects.

The observed excitatory effect of predictive feedback fits the neurophysiologic evidence that feedback is excitatory and the proposed model in the previous chapter. Combining the observed results and previous fMRI evidence, we could have more comprehensive understandings about the roles of predictive feedback and the possible underlying neural circuits.

Chapter III

Another well observed property of neocortex are the oscillations. Neural oscillations are the rhythmic neural activity in the neocortex, which can be observed throughout all levels of activities including spike trains, local field potentials and EEG/MEG.

Recent studies used recordings and micro-stimulation in different layers of neocortex and since superficial and deep layers correspond respectively to the feedback and forward projections (Barbas and Rempel-Clower, 1997; Douglas and Martin, 2004; Wang, 2010), they could tell the frequency of feedback and feedforward oscillations. The evidence suggested that the feedforward and feedback processing have their own signature in frequency: lower frequency (Theta or Alpha frequency) for feedback propagation and higher frequency (Beta or Gamma frequency) for feedforward propagation (Maier et al., 2010; Buffalo et al., 2011; Bastos et al., 2012; van Kerkoerle et al., 2014).

If predictive coding is a universal theory about the feedforward and feedback pathways, it should be able to explain the interactions between the hierarchically higher and lower areas. Thus, the empirical evidence about predictive coding should also show similar oscillatory patterns for feedforward and feedback connections.

Indeed, as we reviewed in the introduction part of this thesis, there are only few evidence about the relationship between oscillations and predictive coding. However, the existing evidence have some problems:

(1) The evidence is not very clear. For example, in the MEG study by Arnal et al, there are several interesting frequency bands in different time points in their correlation between the phase locking factor and ERF, however, they only focused on one frequency band with prior assumption. Furthermore, there is a more significant negative correlation between the theta frequency phase locking and ERF/gamma band power, but they could not tell the possible reasons.

(2) The sources of the oscillations are not clear. In the same study, it is very hard to tell the hierarchical regions of the correlations in their topographical plot.

Thus, even though they proposed a similar oscillation pattern as the recent evidences about feedforward and feedback connections, it is still very hard to convince other researchers that predictive coding actually use their proposed frequency band to communicate.

Here, to further prove that predictive coding is using the same frequency patterns as the observed neurophysiological evidence, we used a similar paradigm as the previous psychophysical study to investigate the relationship between oscillations and predictive coding. We tried to verify the hypothesis that prediction error propagates in a higher frequency and predictive feedback propagates in a lower frequency.

The rhythms of predictive coding: pre-stimulus oscillatory phase modulates the influence of shape perception on luminance judgments

Abstract

Neurophysiological evidence suggests a hierarchy of visual areas pervaded by oscillatory activity. Predictive coding theory provides a canonical neural circuit for the communication between lower- and higher-level areas: a feedforward pathway carrying predictive errors, and a feedback pathway carrying predictions. Because of the iterative nature of this prediction/correction process, we hypothesized that predictions could modulate sensory processing periodically, following the phase of specific brain oscillations. Two gray disks with different versions of the same stimulus, one enabling predictive feedback (a 3D-shape) and one impeding it (random-lines), were simultaneously presented on the left and right of fixation. Human subjects judged the luminance of the two disks while EEG was recorded. We compared the phases of pre-stimulus ongoing oscillations across different post-stimulus judgments. Independently of the spatial response (left/right), the choice of 3D-shape or random-lines as the brighter disk (our measure of the efficiency of predictive coding on each trial) fluctuated along with the pre-stimulus phase of two spontaneous oscillations: a theta oscillation (~5 Hz) in the contralateral frontal electrodes and a beta oscillation (~16 Hz) in the contralateral occipital electrodes. This pattern of results shows that predictive coding takes advantage of higher frequency oscillations in the low-level areas and lower frequency oscillations in the high-level areas. Together with recent studies on predictive coding and feedforward/feedback pathways, our findings support the notion that predictive coding is a periodic process with

faster oscillations in lower areas feeding forward prediction errors, and slower oscillations in higher areas feeding back predictions.

Significance Statement

Predictive coding is an influential model of brain function emphasizing the interactions between feedforward and feedback signals. We investigated the temporal dynamics of predictive coding in the context of shape perception with electroencephalogram (EEG) recordings. By analyzing the relationship between pre-stimulus phase and post-stimulus behavior, we found that contralateral frontal theta-frequency oscillations and contralateral occipital beta-frequency oscillations participated in the predictive coding process, by periodically biasing luminance perception towards the side on which prediction signals were stronger. Together with recent studies, these results support the notion that predictive coding is a rhythmic process, whereby the brain sends feedforward prediction error signals using a faster oscillation, and feedback prediction signals using a slower oscillation.

Introduction

The outside world provides us only the light, but our visual system is capable of extracting the basic features in low-level areas and understanding them as meaningful concepts in high-level areas. Predictive coding theory suggests that the brain employs an efficient coding strategy to achieve this by generating predictions in higher-level areas and comparing them with the incoming sensory signals in the lower-level areas (Rao and Ballard, 1999; Friston, 2005). Previous neuroimaging evidence revealed the existence of such two-way communication (Murray et al., 2002; Harrison et al., 2007; Summerfield et al., 2008; Alink et al., 2010; Egner et al., 2010). However, the underlying mechanisms in this dynamical process, especially in the temporal domain, remain unknown.

It has been proposed that the feed-forward and feedback in predictive coding take advantage of oscillations for information processing (Fontolan et al., 2014). On the one hand, recent neurophysiological evidence on laminar-specific oscillations and the functional roles of different layers suggested a faster oscillation for the feed-forward pathway and a slower oscillation for the feedback pathway (Maier et al., 2010; Buffalo et al., 2011; Bastos et al., 2012; Fontolan et al., 2014; van Kerkoerle et al., 2014). On the other hand, recent studies showed a link between behavioral performance and cortical oscillations in perception (Busch and VanRullen, 2010; Dugué et al., 2011) and reaction time (Drewes and VanRullen, 2011; Song et al., 2014). Since neural oscillations can reflect the cyclic fluctuations of excitability in a network, investigation of the relationship between trial-to-trial variability and the

phase of ongoing oscillations could link specific oscillations to cognitive functions (e.g. attention).

Here, we used this approach to investigate the specific influence of ongoing oscillations on predictive coding, by measuring its effect on perception for different pre-stimulus oscillatory phases. In a typical predictive coding experiment, two conditions must be created: one with strong prediction signals, one without. Since the predictions are sent via feedback signals to lower areas, they will affect the lower-level activity and thus presumably also affect perception. Here, we chose one of the first paradigms in predictive coding to generate different amounts of predictive feedback (Murray et al., 2002): shape perception.

Specifically, 3D-shape outlines and random-lines versions of the same stimuli, similar to the stimuli used in a previous influential study (Murray et al., 2002), were used in this experiment. It has been shown that 3D-shape outlines can be easily recognized and thus produce more predictive feedback than the random-lines versions (Murray et al., 2002). To measure the effect of different amounts of predictive feedback, we asked the subjects to judge the luminance (report the side of the brighter disk) of two gray disks simultaneously displayed on the left and right of fixation on a black background, one containing the 3D-shape outlines and the other containing the random-lines version. The luminance of the disks was adjusted to achieve about 50% choice rate for 3D-shape/random-lines disk (this was achieved by slightly increasing the luminance of the random-lines disk, as demonstrated in one of our previous studies (Han and VanRullen, 2014)). We recorded EEG signals and analyzed the relationship between pre-stimulus oscillation phase and the post-stimulus judgment. We found that, independent from the

spatial choice (left/right side), the phase of 5Hz contralateral frontal and 16Hz contralateral occipital pre-stimulus oscillations modulated the subject's choice of a brighter 3D-shape disk (more effective predictive feedback) or a brighter random-lines disk (less effective predictive feedback). Since higher hierarchical level areas are assumed to send predictive feedback and lower hierarchical level areas to send predictive error, our results imply that the brain sends predictive feedback periodically at a preferred phase of a theta frequency oscillation in the frontal region, and sends predictive errors periodically at a preferred phase of a beta frequency oscillation in the occipital region.

Results

Human observers judged the luminance of two disks that were presented for 150ms on the left and right of a central fixation point. The disks contained different versions of the same stimulus, one with a 3D-shape enabling predictive feedback, and the other with a random-lines version of the same shape which impeded predictive feedback (Figure 4-1). Before the stimulus onset there was a random period of time (1000-1500ms) with only the fixation point on the screen. After the stimulus offset a question mark appeared at the center and the subjects were instructed to report the side with the brighter disk. In the main experimental trials, the luminance of the disks was adjusted, such that observers reported the 3D-shape disk as brighter in half of the trials. 15% of trials were catch trials: extreme luminance values were assigned to one disk to monitor the subject's ability to judge the luminance difference.

Behavioral Results

On average, subjects judged the 3D-shape disk as brighter in half of the trials (49.24% ± 1.57%, mean ± standard error of the mean, SEM) in the main experimental condition, as expected. The luminance judgment correct rate in the catch trials (subjects judged the disk with higher luminance value as brighter or judged the disk with lower luminance value as darker) was high (93.98% ± 1.83%, mean ± SEM), indicating that subjects were adequately engaged in the luminance judgment task.

Electrophysiological Results

We focused on the relationship between oscillatory phase and the trial-by-trial variations in the efficiency of predictive coding. EEG was recorded during the experiment. We expected the relation between oscillatory phase and behavior to be most visible in the pre-stimulus time window, where phase information reflects spontaneous fluctuations in neuronal excitability (Bishop, 1932; Buzsáki and Draguhn, 2004; Fries et al., 2007; Busch et al., 2009; Jensen et al., 2012). In contrast, post-stimulus phase information is driven to a large extent by stimulus-locked activity (e.g. evoked potentials) and is thus further removed from spontaneous activity. We used classical stimuli (Murray et al., 2002) for inducing different amounts of predictive feedback on the left and right of the screen: 3D-shape and random-lines versions of the same stimuli (Figure 4-1; the 3D-shape version enabling predictive feedback, the random-lines simultaneously impeding it). To measure the effective amount of predictive feedback on each trial, we probed the perceived luminance of the disks under the stimuli. We have previously demonstrated that the net effect of predictive feedback on these stimuli is a relative increase

of perceived luminance for the disk containing the 3D-shape (Han and VanRullen, 2014). Here, this net effect was compensated on each trial by slightly lowering the luminance of that disk so that the average likelihood of perceiving either disk brighter was about 50% (see Methods); therefore, residual fluctuations of luminance perception on every trial can be thought to arise from trial-by-trial fluctuations in the efficiency of predictive coding (the 3D-shape disk may still be perceived brighter on trials where predictive coding was more efficient than average, and darker on trials where it was less efficient than average). Of course, spatial bias and/or trial-by-trial fluctuations in the direction of spatial attention may well also contribute to the choice of which disk appears brighter on a given trial. Thus, for each subject we divided all trials into two datasets based on their spatial choice (left-side choice vs. right-side choice), and we only investigated the relation between pre-stimulus EEG phase and 3D-shape/random-lines choice within each dataset. As the correlates of choosing the prediction-consistent stimulus (3D-shape) were expected to be strongest on electrodes contralateral to that stimulus, which would map onto opposite hemispheres for the two datasets, before plotting any scalp topographies we permuted the electrode positions (symmetrically across the midline axis) of all right-side choice trials. This procedure resulted in a mapping of ipsilateral effects to the spatial choice onto left electrodes, and contralateral effects onto right electrodes.

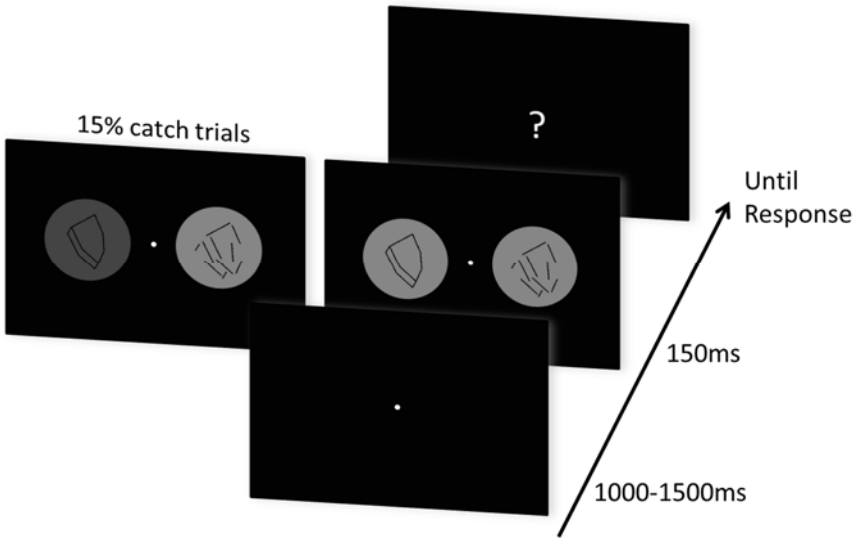


Figure 4-1 Illustration of the experimental paradigm. In each run of trials, a blank screen with only a central fixation point was presented for 1000 to 1500ms randomly. Then, two circular gray disks, one with a 3D-shape stimulus in the center and the other with random-lines, were presented randomly on either side (left or right) of the fixation point for 150ms. Subsequently, a question mark appeared in the center of the screen. Subjects were instructed to fixate the fixation point all the time, and report the side of the brighter disk with the corresponding arrow key after the question mark appeared. In the main experimental trials, luminance values of the disks were adjusted to obtain a 50% selection probability of 3D-shape/random-lines disk. In addition, there were 15% catch trials intermixed with the main experimental trials to monitor the subjects' ability to judge the luminance difference throughout the experiment. In these catch trials, one disk was 20% brighter/darker than in the main experimental trials.

We estimated the relation between EEG phase and predictive coding via the phase opposition product (POP, see Methods). This measure should be maximal when 3D-shape choice trials (prediction-consistent) and random-lines choice trials (prediction-inconsistent) tend to have opposite phase values. For each subject, dataset, electrode, time point

and oscillatory frequency, we obtained surrogate POP values (80,000,000 surrogates) by randomly permuting the trial outcomes, keeping the number of trials constant. Both real and surrogate POP values were averaged across datasets and subjects. The significance was determined as the proportion of surrogate POP values that were more extreme than the observed value. P-values were corrected for multiple comparisons across time points, frequencies and electrodes (100×30×64) using the FDR method (FDR α=0.05, corresponding to a P value threshold of 9.53×10^{-6}). To show the overall POP in the time-frequency domain, a z-score was computed by comparing the real POP values (combined across all subjects, datasets, and electrodes) to the mean and standard deviation of a null-hypothesis distribution with 10,000 surrogate POP values (generated using the same procedure described before, and also combined across all subjects, datasets, and electrodes). This analysis revealed a significant relation between the post-stimulus 3D-shape/random-lines choice and two pre-stimulus oscillations (Figure 4-2. A): one theta-frequency oscillation (~3.1 Hz to 7.6 Hz) in the time window from -545 ms to -268 ms, and one beta-frequency oscillation (~13.2 Hz to 25.7 Hz) in the time window from -107 ms to -25 ms. Green outlines mark the significant time-frequency regions (at least one significant electrode) after FDR correction.

Scalp topographies of the z-score show that the two oscillatory effects involve distinct electrode groups and presumably distinct brain regions (Figure 4-2. B and C): the theta-frequency effect is maximal over frontal regions and the beta-frequency effect over occipital regions. In both cases, these effects are contralateral to the side that subjects chose as "brighter" (i.e., the right side of the topographies, due to our electrode permutation procedure). Electrodes with at least one significant time-

frequency point (after FDR correction) inside the corresponding time-frequency window are highlighted in green.

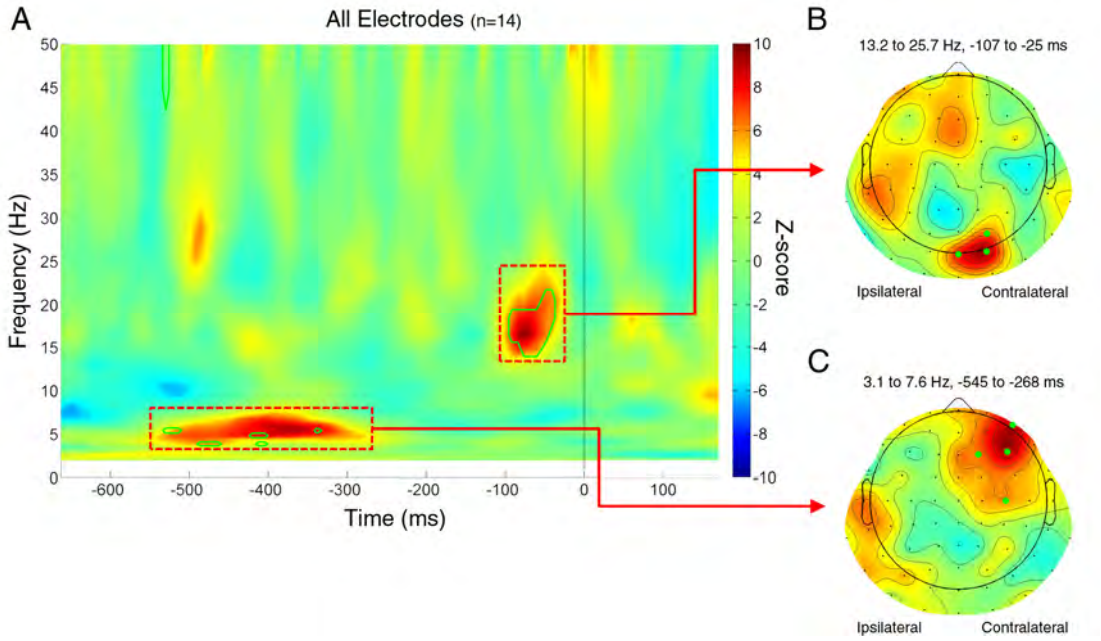


Figure 4-2 Pre-stimulus EEG phase predicts luminance judgment of 3D-shape disk vs. random-lines disk. (A) The relation between pre-stimulus phase and luminance judgment (our measure of predictive coding) is assessed using POP values (phase opposition product; see details in Methods). The time-frequency map is the z-score of observed POP values (combined across all subjects, datasets and electrodes), each value compared with a null-hypothesis distribution of 10,000 surrogate POP values (also combined across all subjects, datasets and electrodes) characterized by its mean and SD. Time 0 indicates stimulus onset. P-values were derived from a comparison of POP values against 80,000,000 surrogates, and corrected for multiple comparisons across all time points, frequencies and electrodes using the FDR method (FDR a = 0.05, corresponding to a P value threshold of 9.53×10^{-6}). The green outlines mark the significant time-frequency regions (at least one significant electrode) after FDR correction. A significant relation is apparent between the effect of shape perception on luminance judgments and the EEG phase of ~5 Hz and ~16 Hz

pre-stimulus oscillations. (B) Scalp topography of (z-scored) POP values around 16 Hz (frequency range 13.2 to 25.7 Hz; time range -107 to -25 ms). Electrodes highlighted in green have at least one significant time-frequency point (after FDR correction) inside the corresponding red box. Due to our electrode-permutation procedure, in this topography the electrodes ipsilateral to the spatial choice are displayed on the left and those contralateral to the spatial choice are displayed on the right. Thus, the topography shows a contralateral occipital effect for the 16 Hz oscillation. (C) Same as B, but for the ~ 5 Hz oscillations (frequency range 3.1 to 7.6 Hz; time range -545 to -268 ms). The topography shows a contralateral frontal effect for the 5 Hz oscillation.

To quantify the influence of pre-stimulus oscillations on post-stimulus choice, we binned single trials according to the phase at the optimal time-frequency point (for the theta oscillation: -397 ms, 5.4Hz; for the beta oscillation: -68 ms, 16.5 Hz). Single trials were thus sorted in 13 phase bins based on the average phase of the significant electrodes for each oscillation (four frontal electrodes for the theta oscillation, three occipital electrodes for the beta oscillation). For each phase bin we then computed the post-stimulus choice probability of the 3D-shape disk. These choice probabilities were normalized by dividing them by the overall 3D-shape choice probability across all phase bins. For each experimental dataset (left- vs. right-side choice), phase bins were rotated such that the phase at which 3D-shape disk choice probability was largest was aligned to a phase angle of zero. As a result of this alignment, the 3D-shape choice probability is necessarily maximal at a phase angle of zero; therefore, the zero-phase bin was discarded from further analyses. For both frequencies, the 3D-shape disk choice probability monotonically decreased to a minimum at the opposite phase angle, confirming that pre-stimulus phase affected post-stimulus judgment (Figure 4-3). A one-way ANOVA showed that both pre-stimulus

theta phase and pre-stimulus beta phase significantly modulated the 3D-shape disk choice probability (for theta oscillation, $F_{(11, 27)} = 3.95$, $p = 2.23 \times 10^{-5}$; for beta oscillation $F_{(11, 27)}=6.17$, $p=3.86 \times 10^{-9}$). The magnitude of each effect was determined as the difference between the maximum and minimum 3D-shape disk choice probabilities across all phase bins. The frontal theta oscillation accounted for a difference of ~14% of the 3D-shape disk choice probability between phase bins, and the occipital beta oscillation accounted for a difference of ~19%.

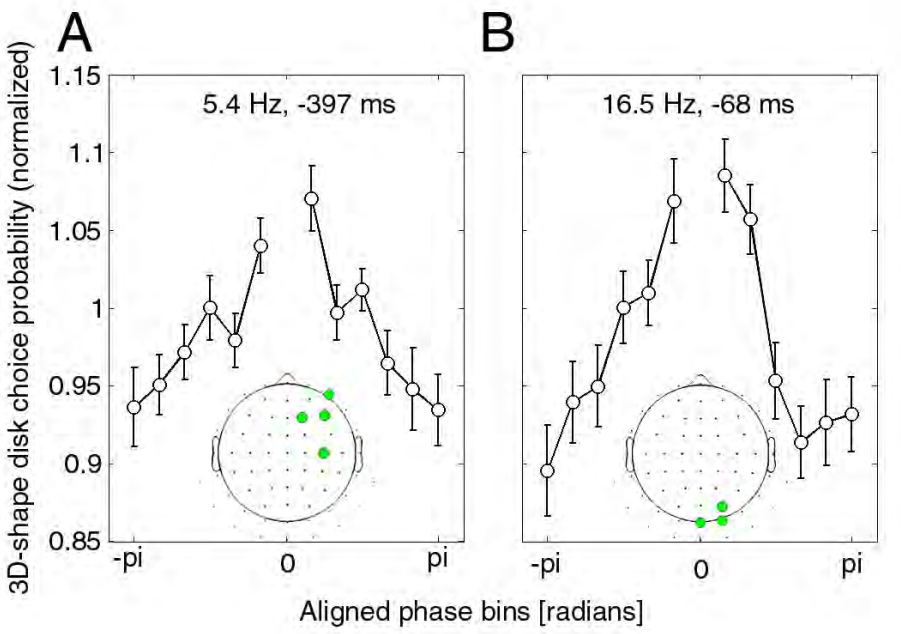


Figure 4-3 Normalized choice probability of 3D-shape disks as a function of pre-stimulus phase. (A) Relationship between frontal pre-stimulus theta phase and choice of 3D-shape disk as the brighter disk. Single trials were binned into 13 bins, centered on the maximal phase bin for each subject (central bin was then discarded). The curve indicates that the oscillatory phase of frontal electrodes (4 significant electrodes shown in inset topography) at 5.4 Hz and -397 ms modulates the luminance judgment by ~14%. Error bars represent SEM across subjects. (B) Same as A, but for the occipital beta pre-stimulus phase. The oscillatory phase of occipital electrodes (3 significant electrodes shown in inset

topography) at 16.5 Hz and -68 ms modulates the luminance judgment by ~19%.

Because the time-frequency analysis relies on signal convolution with wavelet filters whose duration is non-negligible, one might wonder whether the observed pre-stimulus phase differences could actually be driven by stimulus-evoked activity. For example, at 16 Hz the above time-frequency analysis used a 250 ms time window (4 cycles, 125 ms from the past and 125 ms into the future); thus, significant phase effects observed at -67ms pre-stimulus may be contaminated by post-stimulus activity. To rule out such contamination, we repeated the POP time-frequency analysis with one-cycle wavelets at all frequencies, and compared the timing of pre-stimulus phase effects with the time-frequency region of possible post-stimulus contamination, determined using the wavelet window length at each frequency (Figure 4-4). Both theta- and beta-frequency phase effects were replicated in this analysis, and were found to lie outside of the possible post-stimulus contamination zone.

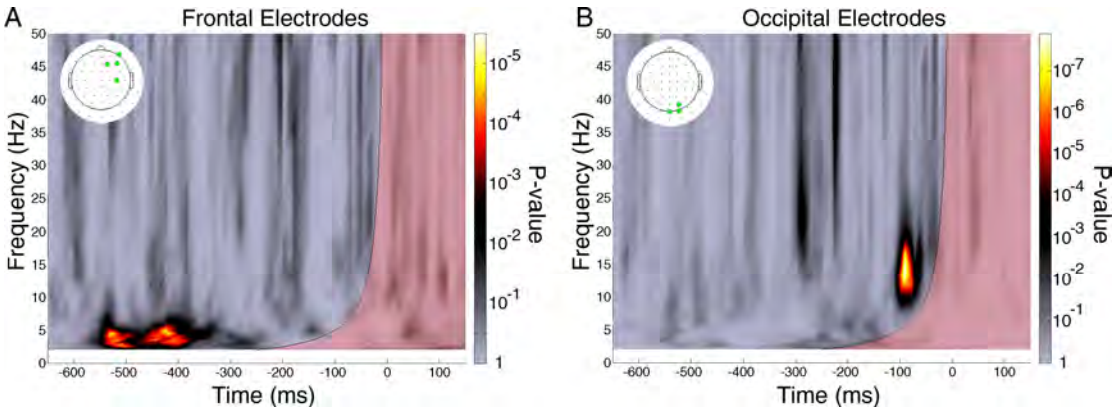


Figure 4-4 Significance of POP values in a one-cycle wavelet analysis. (A) P-value map of the POP values combined across previously identified frontal electrodes (green points in inset topography) for the theta-frequency phase effect. The P-values were calculated by comparing the observed POP values with 80,000,000 surrogates. The semi-transparent red area on the time-frequency map indicates the zone of possible contamination by post-stimulus activity (based on the wavelet window length at each frequency, centered on the time of stimulus onset, 0 ms). The previously observed theta-frequency phase effect lies outside of the contamination zone. (B) Same as A, but for the beta-frequency phase effect. The beta-frequency phase effect also lies outside the contamination zone. Altogether, these findings indicate that pre-stimulus phase differences are not caused by post-stimulus evoked activity.

We also ascertained that phase effects were not caused by any eye movement artifacts that may have survived our artifact rejection procedure. For example, the observed pre-stimulus phase differences could be thought to reflect different patterns of eye blink or saccades for different perceptual outcomes. Therefore, we applied our POP time-frequency analysis to the horizontal and vertical EOG signals. P-value maps (obtained by comparison of POP values against 80,000,000 surrogates) did not reveal any signs of systematic eye movements in

either the theta- or the beta-frequency bands (Figure 4-5), ruling out an explanation of our pre-stimulus phase effects in terms of ocular artifacts.

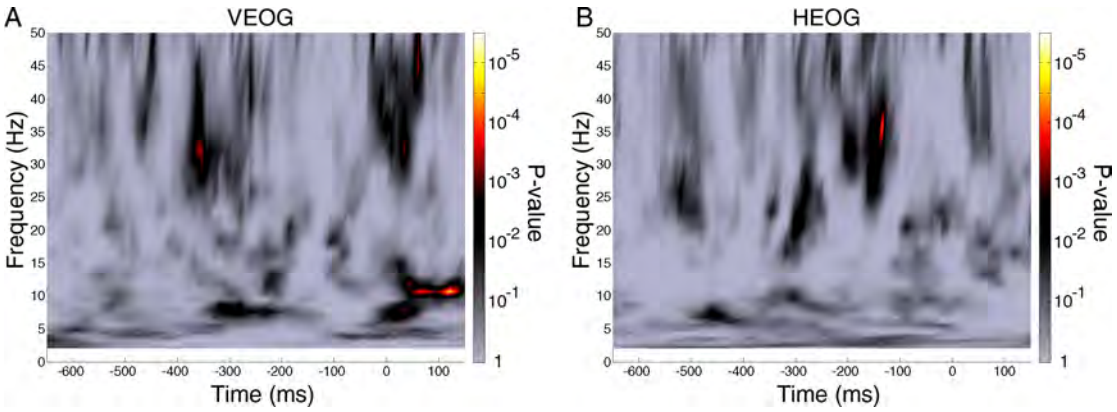


Figure 4-5 Significance of the POP values of VEOG and HEOG signals. (A) P-value map of the POP values for the VEOG. The P-values were calculated using a similar procedure as in the main analysis: comparing the observed POP values with 80,000,000 surrogate POP values. (B) Same as A, but for the HEOG. There was no significant pre-stimulus time-frequency window with significant POP values in either VEOG or HEOG, indicating that the observed phase effects were not due to ocular artifacts.

Discussion

We investigated the temporal dynamics of predictive coding by exploring the relation between pre-stimulus oscillatory phase and the presumed trial-by-trial variations in predictive feedback. We used 3D-shape outlines and random-lines versions of the same stimuli (as in one of the seminal predictive coding studies (Murray et al., 2002)) to induce different amounts of predictive feedback (Figure 4-1), and measured the corresponding effects on luminance judgment as trial-by-trial markers of the efficiency of predictive coding. We found that two pre-

stimulus ongoing oscillations from different regions and frequencies could strongly influence the luminance judgment: a contralateral frontal theta oscillation and a contralateral occipital beta oscillation (Figure 4-2). The phase of the theta oscillation before stimulus onset could explain 14% of the luminance judgment difference while the phase of the beta oscillation could explain 19% (Figure 4-3). Control analyses ruled out contamination of the phase-behavior relationship by post-stimulus activity (Figure 4-4) or ocular artifacts (Figure 4-5). These results not only imply that predictive coding is a periodic process, but also reveal two periodicities with different sources. Since the occipital and frontal signals likely reflect activity from hierarchically lower and higher areas, respectively, and since predictive coding theory suggests that the brain sends back predictions from higher areas and sends predictive errors from lower areas, our results suggest one possible temporal dynamic for predictive coding: predictions sent periodically at a theta frequency, predictive errors sent periodically at a beta frequency.

The experimental paradigm used in this study takes advantage of the relationship between shape perception and predictive coding: 3D-shape outlines are assumed to generate more predictive feedback than the random-lines version of the same stimulus. Murray et al (2002) used similar stimuli to provide one of the first evidence of predictive coding: compared to the random-lines, 3D-shape outlines increased activity in the lateral occipital complex (LOC), but decreased it in primary visual cortex (V1), suggesting an increase of predictive feedback accompanied by a decrease in prediction errors (Murray et al., 2002; Clark, 2013). Here, we used the same paired stimuli as in the original study, placed them on two gray disks and asked subjects to judge the luminance of the disks. This luminance judgment, associated with

perceived contrast, is likely to have a positive and monotonic relationship with neural activity in early visual cortex (Dean, 1981; Albrecht and Hamilton, 1982; Goodyear and Menon, 1998; Boynton et al., 1999). Thus, the variability in luminance judgment associated with the 3D-shape vs. random-lines stimuli could reflect trial-by-trial changes in the effect of predictive feedback on neural activity in early visual cortex.

Previous fMRI studies showed that shape perception could not only reduce (Murray et al., 2002), but also up-regulate neural activity in V1 (Kok and de Lange, 2014). Our own previous study found that 3D-shape disks were generally perceived brighter than random-lines disks, and that this effect could be attributed to predictive coding rather than attentional biases (Han and VanRullen, 2014). In the present study, we compensated for this net effect by adjusting the disks' luminance to obtain a 50% selection probability of 3D-shape/random-lines disks, and we focused on the remaining variability in luminance judgement as a trial-by-trial marker of the efficiency of predictive coding. On the other hand, systematic spatial biases (e.g. a general tendency to respond to the left or right stimulus) and/or trial-by-trial fluctuations in the direction of spatial attention can also be expected to affect the luminance judgement (Carrasco et al., 2004). As a matter of fact, spatial attention itself appears to involve a periodic process (Busch et al., 2009; Busch and VanRullen, 2010; Landau and Fries, 2012; Fiebelkorn et al., 2013) which could potentially influence the luminance judgement. We carefully avoided these potential confounding factors by dividing the trials into two datasets based on the post-stimulus spatial response (left/right) and performing the analysis within each dataset. If spatial attention biases, for example, were the only cause of the perceived luminance changes, the left-response dataset would pool all trials with a left-side attention

bias (and similarly for the right-response dataset), and within each dataset pre-stimulus oscillatory phases would not bear any relation to post-stimulus luminance judgments. The existence of significant phase-behavior relationships in our analysis can therefore be safely attributed to predictive coding mechanisms rather than spatial attention or other biases.

Neurophysiological recordings have shown that feedforward and feedback may take advantage of oscillations in different frequencies. Laminar recordings showed that high-frequency oscillations are prominently generated in superficial layers and low-frequency oscillations in deep layers (Roopun et al., 2006; Maier et al., 2010; Buffalo et al., 2011). Since superficial and deep layers correspond respectively to the main sources of feedback and forward projections (Barbas and Rempel-Clower, 1997; Douglas and Martin, 2004; Wang, 2010), it follows that feedforward communication takes advantage of high-frequency oscillations and feedback takes advantage of low-frequency oscillations. A recent study with simultaneous recordings and micro-stimulation in different layers in V1 and V4 confirmed this notion (van Kerkoerle et al., 2014). Several authors have independently proposed that low-frequency oscillations send predictions via feedback, while high-frequency oscillations send predictive errors via feedforward (Todorovic et al., 2011; Arnal and Giraud, 2012; Bastos et al., 2012; Yordanova et al., 2012; Bauer et al., 2014; Fontolan et al., 2014). Our results provide clear support for this hypothesis at the EEG and behavioral level.

Our results also provide supportive evidence for the hypothesized functions of frontal theta-band and occipital beta-band oscillations. Our

conclusions are in line with the notion that 5-10 Hz oscillations could contribute to “top-down” control (Jensen et al., 2012; VanRullen, 2013), which has already been suggested based on attentional phase effects on perception (Busch et al., 2009; Busch and VanRullen, 2010), reaction time (Drewes and VanRullen, 2011; Song et al., 2014; Huang et al., 2015) and perceptual variability in TMS-induced effects (Dugué et al., 2011, 2015). We found the origin of such theta periodicity in contralateral frontal electrodes, compatible with the involvement of frontal areas in the top-down controlling process (Summerfield et al., 2006; Summerfield and Egner, 2009; Summerfield and de Lange, 2014) and with the involvement of 5-10 Hz oscillations in this region (Phillips et al., 2014). On the other hand, local field potential (LFP) recordings showed that, in mammalian visual cortex, beta frequency oscillations are also prominent during the deployment of top-down control (Lopes da Silva et al., 1970; Bekisz and Wróbel, 2003; Buschman and Miller, 2007; Bosman et al., 2012; Grothe et al., 2012). Our findings of beta frequency phase effects on predictive feedback in the occipital area are concordant with such LFP results and suggest a valuable role for the beta frequency oscillations in predictive coding.

In summary, we measured the relation between pre-stimulus oscillations and a predictive feedback-induced effect to investigate the neural oscillations involved in predictive coding. We found that the pre-stimulus phases of frontal theta-frequency oscillations and occipital beta-frequency oscillations jointly determine post-stimulus subjective judgments. These results shed light on the temporal dynamics of predictive coding, and suggest a periodic predictive coding process with faster oscillations in lower areas and slower oscillations in higher areas.

Materials and Methods

Subjects

Fifteen volunteers participated in the experiment. One participant was excluded from the analysis due to the poor behavioral performance in catch trials (<60% trials were correctly reported, with a chance level of 50%, see below). Fourteen participants remained in the sample (8 female, mean age 28.01 ± 4.81 years, four left-handed, four with left eye dominance). All subjects had normal or corrected to normal vision.

Apparatus

Stimuli were presented at 57 cm distance using a desktop computer (2.09 GHz Intel processor, Windows XP) with a cathode ray monitor (resolution: 800×600 pixels; refresh rate: 140 Hz, Gamma corrected luminance function). Stimuli were designed and presented via the Psychophysics Toolbox (Brainard, 1997) running in MATLAB (MathWorks).

Stimuli and tasks

Stimuli consisted of a central white fixation point (diameter: 0.2 degrees of visual angle) and two circular gray disks (diameter: 4 degrees each) presented randomly to the left and right of fixation (3 degrees eccentricity). One 3D-shape stimulus was in the center of one disk (3D-shape disk) and one random-lines version of the same stimulus in the other (random-lines disk). The 3D-shape and random-lines stimulus pair

was randomly chosen from twenty pairs of stimuli generated beforehand using a method similar to Murray et al. (2002): 3D-shapes were generated by randomly selecting 4–6 vertices, connecting the vertices and adding small extensions to render perceived depth; random-lines stimuli were created by breaking the 3D-shape at its intersections and randomly shifting the lines within the display (Murray et al., 2002). The diameter of both 3D-shape and random-lines stimuli was 3 degrees. The stimulus outlines were black.

Before stimulus onset, there was a blank screen with only the fixation point that lasted from 1000 to 1500ms (random uniform distribution). Then the two disks and the fixation point appeared for 150ms. After that, a question mark appeared in the center of the screen. There were two kinds of randomly mixed experimental trials: the main experimental trials and the catch trials. In main experimental trials, the luminance of the disks was adjusted (i.e. the random-line disks were set 1.45% brighter than the 3D-shape disks) based on a previous study (Han and VanRullen, 2014) to obtain an average 50% selection rate of 3D-shape/random-lines disks. In catch trials, one of the disks had its luminance value changed up or down by 20% compared to the luminance used in the main experimental trials, while the other disk kept the same luminance as in the main experimental trials. Subjects were presented with 4 or 8 blocks of 200 trials with 85% main experimental trials and 15% catch trials (the first 6 of the 14 subjects performed only 4 blocks of the present experiment, together with 4 blocks of another experiment that was eventually canceled and whose data were not analyzed). Subjects were instructed to fixate the fixation point all the time, judge the luminance of the disks and respond using the arrow keys (left arrow to indicate that left disk is brighter, right arrow for right disk brighter) on a

standard 105 key keyboard when the question mark appeared. There was no feedback after the response.

EEG data acquisition and analysis

EEG was recorded at 1024 Hz using a Biosemi system (64 active electrodes). Horizontal and vertical electro-oculograms (EOG) were recorded by three additional electrodes around the subjects' eyes. For data pre-processing, the EEG and EOG data were downsampled offline to 256 Hz, re-referenced to average reference and epoched around the stimulus onset in each trial for data analysis *via* the MATLAB (MathWorks) and EEGLAB toolbox (Delorme and Makeig, 2004). Individual electrode data were visually inspected, and channel data containing artifacts were interpolated by the mean of adjacent electrodes (three subjects had one electrode containing artifacts, one subject had two; the positions of the interpolated electrodes were different across subjects).

As the post-stimulus spatial choice was lateralized on each trial to the left or right side, the pre-stimulus oscillatory correlates of the post-stimulus luminance judgment may not only reflect the oscillation's influence on shape perception and predictive coding, but also its influence on spatial choice (i.e., pre-stimulus oscillations may bias the left/right spatial choice independently of the 3D-shape/random-lines content inside of the disk). To avoid any contribution from the spatial choice, we first divided the trials for each subject into two trial datasets based on the post-stimulus spatial choice, and performed the time-frequency analysis (described below) within each dataset. We reasoned that this analysis would lead to shape perception correlates not on a given fixed set of electrodes, but rather on different electrode groups depending on the side of

choice (i.e., electrodes "contralateral" or "ipsilateral" to the spatial choice). Therefore, we arbitrarily chose to permute the electrode locations for the dataset corresponding to a right-side choice: we replaced the left-hemisphere electrodes by the symmetric ones from the right and vice versa (midline electrodes were unaffected). With this new electrode assignment, left-hemisphere electrodes would thus always correspond to those ipsilateral to the spatial choice, and right-hemisphere electrodes to contralateral ones.

For the time-frequency analysis, time-frequency transformations were first generated over all channels using EEGLAB with a function akin to a wavelet transform, starting with 3 cycles at 2Hz and increasing to 5 cycles at 50 Hz in the multiple-cycle analysis, and with 1 cycle from 2Hz to 50 Hz in the one-cycle analysis. This yields a complex representation of the amplitude, A, and the phase, φ, for trial j at time t and frequency f:

$$\mathrm{tf}_j(t,f) = A_j(t,f)e^{i\varphi_j(t,f)}$$

The phase of this representation can be extracted by normalizing the complex vector to the unit length:

$$\Phi_j(t,f) = \frac{\mathrm{tf}_j(t,f)}{\left|\mathrm{tf}_j(t,f)\right|}$$

Phase locking value (PLV or inter-trial coherence, ITC) measures the phase consistency across trials. We calculated the PLV using the method described previously (Lachaux et al., 1999):

$$PLV(t,f) = \frac{1}{N}\left|\sum_{j=1}^{N} \Phi_j(t,f)\right|$$

where N is the number of trials in one group of trials.

Here, we wanted to evaluate the relation between the pre-stimulus oscillatory phase and the influence of shape perception on luminance judgment (our measure of the efficiency of predictive coding). Would a particular pre-stimulus phase occur more frequently for trials with post-stimulus 3D-shape disk choice, and the opposite phase for trials with post-stimulus random-lines disk choice? In the pre-stimulus period, because intertrial intervals are randomized and unpredictable, the phase of the spontaneous EEG signal at a given pre-stimulus time should follow a uniform distribution over all trials. However, if there is a systematic relation between EEG phase and behavioral outcome, higher-than-chance phase-locking should be observed in each of the trial subgroups. In that case, the product or the sum of the two subgroup phase-locking values could summarize, in a single variable, the strength of the phase-behavior relation (Busch et al., 2009; VanRullen et al., 2011). Here we thus introduce a new measure of the phase-behavior relation: *Phase Opposition Product (POP)*. This measure is calculated using the product of the phase locking values of different trial subgroups:

$$POP = PLV_a \cdot PLV_b$$

To accurately assess the significance of the phase-behavior relation without any assumption about the probability distribution of the POP values, we performed a nonparametric permutation test: We first computed the POP values for each point in the time-frequency plane

from -650 to 150ms, from 2 to 50 Hz for each electrode, dataset, and subject and then averaged across all datasets and all subjects. Surrogate POP values were obtained by randomly assigning the trials to one or the other condition for each subject (keeping the number of trials in each condition constant) and recalculating the grand-average POP values. We computed the P value by simply counting the number of surrogate POP values that were more extreme than the observed value. Here, we used 80,000,000 surrogates and thus assigned the P value of 1.25×10^{-8} to the points without any more extreme POP values in the surrogates. The P values were corrected for multiple comparisons over time points, frequencies and electrodes using the FDR method (FDR α=0.05, corresponding to a P value threshold of 9.53×10^{-6}). To show the overall POP in the time-frequency domain, we computed a z-score by combining the observed POPs across all datasets, subjects, and electrodes and comparing the value with the mean and SD of a null-hypothesis distribution with 10,000 surrogate POP values (generated using the procedure described before, and also combined across all electrodes, subjects, and datasets).

Conclusion

In this chapter, we investigated the relationship between the predictive coding and neural oscillations. By using a similar paradigm as in the previous chapter, we analyzed the relationship between pre-stimulus phase in the recorded EEG signals and post-stimulus behavior (choosing the 3D-shape disk or random-lines disk as the brighter disk, or predictive feedback's efficiency on perception). The results showed that a theta oscillation (~5 Hz) in the contralateral frontal electrodes and a beta oscillation (~16 Hz) in the contralateral occipital electrodes are correlated with the efficiency of predictive coding on each trial. Control analysis exclude stimulus-evoked activity contamination and eye movement artifacts as the alterative explanations of the observations.

Our results support the notion that predictive coding is a periodic process with faster oscillations in lower areas feeding forward prediction errors, and slower oscillations in higher areas feeding back predictions which has been proposed by other researchers. However, this study provided much better evidence than the previous studies:

(1) We showed a very clear evidence of two oscillations with different frequencies which their pre-stimulus phase can modulate the post-stimulus luminance judgement. The two oscillations are the only significant oscillations in the pre-stimulus time window.
(2) The topographical plot showed very clear sources of the different oscillations: contralateral frontal electrodes for the ~5 Hz oscillation and contralateral occipital electrodes for the ~16 Hz oscillation. The positons of the electrodes (frontal and occipital) are clear indications

of the hierarchically higher or lower area activity (frontal for higher, occipital for lower).

(3) The 80,000,000 simulations guaranteed the statistical accuracy of our analysis. Since the measures of EEG usually do not show normal distributions in statistic, our non-parametric method ensured that the observed results are genuine.

Together, we provided clear and convincing evidence for the relationship between the predictive coding and oscillations. We supported the view that oscillations with different frequencies may have a different role in predictive coding.

Discussion

Summing-up

Motivation

In this thesis, we investigated predictive coding and its relationship with perception and oscillations. Predictive coding is a promising theory of the brain and many researchers have an interest in it. However, there are a thousand Hamlets in a thousand people's eyes. The ways that the researcher treat predictive coding theory is determined by his/her background and experience in the field of research.

For example, many cognitive neuroscientists treat predictive coding as a phenomenon. In their points of view, predictive coding is an instrument that can explain the inhibitory effect in the observation. Thus, there are numerous studies trying to find out an inhibitory feedback effect and connect the known inhibitory effects (such as repetition suppression, mismatch negativity) with predictive coding. There is no system difference between the predictive coding and the magical attention effect: it is only a way to explain the observed data. Thus, predictive coding can appear in the same data with attention in a parallel way: explaining the excitatory effect using attention and explaining the inhibitory effect using predictive coding.

On the other hand, many theoretical neuroscientists treat predictive coding as one of the Bayesian approaches to brain function and there is no difference between predictive coding and Kalman filter: they are

all factors in Bayesian equations. In this situation, the researchers consider predictive coding in a very abstract way, so that all the neurophysiological limitations are not applying to predictive coding. Sometimes, different types of neurons, different layers of neocortex or different types of neural connections set to match the different components of predictive coding while the proposed model is simply impossible under well recognized principles of the brain.

For sure, these ways of research indeed helped us to believe that predictive coding really exists in the brain and it has a functional significance in the computation. However, in my mind, both of the methods of treating predictive coding are not optimal and we can learn very limited information about the brain from these experiments.

In this thesis, we applied a different perspective: keeping the core of the predictive coding principle unchanged and trying to fit the model as close as possible to the neurophysiological evidence since the neuroanatomy is very strong and stable and there are few alternative ways to explain neurophysiological evidence. I believe that if predictive coding is the universal principle for the interaction between hierarchically higher and lower areas, we can find out a special design in our brain and this design is already lying under the well observed evidence of the brain. This design may not have the exact appearance as described in the original predictive coding model, but it must share a common working principle as the predictive coding. If we can find out this unique design, we can develop a "better model" for predictive coding and we can understand more about the brain in the process. Then, we can apply the "better model" to behavioral experiments and more neurophysiological experiments. If we can keep on doing this, we

cannot only verify the model itself, but also integrate the existing knowledge about the brain and finally build an ultimate model of the brain.

The content

The first step to accomplish this goal is to identify the neurophysiological evidence that we believe to be fundamental and universal. Even though neuroscience is still a young field of research, we do know some facts of the brain for sure. For instance, we know that the basic elements of our working brain are the neurons and they have the physical axons and use the action potential to carry the information from one neuron to another. If we believe these as the facts of the brain, we can eliminate lots of alternative explanations of observed data.

Thus, in the first part of my thesis, I reviewed neuroscience facts as the limitations that we are required to obey in the model and see them as the foundation of the principle of the brain. In this part, I reviewed my current understanding about neuron and neocortex.

First, I reviewed our knowledge of the neuron from a historical point of view. From Ramón y Cajal, the first scientist that reported neurons as individual (1888), to the different neurons discovered by the pioneers of the field of neuroscience, we could understand the origins of basic ideas in neuroscience. Then I reviewed the known features of the neurons: three categories of the properties were investigated: physical properties, neurotransmitters and electrophysiology.

For the physical properties, I first reviewed the knowledge about the neuronal shape related information such as the difference between the Pyramidal neurons and stellate neurons, the size of the neuron, the size of the different parts of the neuron and the numbers of different types of neurons. With clear photos of the neurons in the neocortex, we could know directly about the neurons. The shapes of neurons alone could tell us a lot about the limitations of the neuronal models. For example, the axonal field of the stellate neurons only have a length of 100-150 micrometers, which suggested that it is impossible for stellate neurons to send information to other areas in the brain, and the pyramidal neuron are the only known neuron type that can send information to another area in the brain. The absence of layer 4 stellate neurons in non-sensory regions also suggested a weaker role for the stellate neurons. Then, I reviewed the information about the dendritic spine, which could be only found in the excitatory neurons. From this information and the well accepted idea that dendritic spines are the key elements for the spike-timing dependent plasticity and the building of long-term potential and long-term depression, it seems to be obvious that it would be hard to learn the synaptic weights to the inhibitory neurons.

Then, I reviewed the two types of neurotransmitters in the brain and the two types of neurons in the brain: glutamate-releasing neurons (excitatory) and GABA-releasing neurons (inhibitory). Dale's law also forbids the neurons to be both glutamate-releasing and GABA-releasing. Thus, the excitatory and inhibitory neurons are the basic types of neurons that we are interested in. Further information about the excitatory and inhibitory synapses tells us that most of the selectivity are built-in only the pyramidal neurons (since most excitatory synapses land on the dendrites shafts of spiny stellate neurons).

At last, I reviewed electrophysiological properties of the neurons. In this part, since there are plenty of different kinds of electrophysiological properties, I used a modern data collection method to gain an accurate understanding of these properties: I took the data from 64 studies and used the statistical values as my understanding. For all the electrophysiological properties, we found out that some of the values are always similar across different types of neurons such as the resting membrane potential, the spike threshold and spike amplitude. Some other values are different across different types of neurons such as the input resistance, membrane time constant, spike width (excitatory neurons usually have a bigger spike width, the Martinotti cell is an exception), firing rate (inhibitory neurons usually fires faster except the Martinotti cell), Ahp Amplitude and adaptation ratio (inhibitory neurons usually are easier for adaptation except Martinotti cell).

Secondly, I reviewed my understanding of the neocortex. Neocortex is the center of visual perception, auditory perception, motor controlling, reasoning, language and conscious thought. It is one of the most important parts of the brain since most of the functions of the brain depend on it. I reviewed my understanding about the structure, the connectivity and the temporal dynamic of neocortex. Then I combined these understandings and other neural circuits information and proposed a canonical neural circuits.

To understand the structure of neocortex, I make good use of our knowledge about the two kinds of levels to investigate that: the macro-structure (different areas) and the micro-structure (different layers). Since our current understanding about the structure of neocortex is based on the studies made more than 100 years ago (e.g. Works by

Brodmann), we used a historical point of view of the evolution of our understanding of the structure. For the macro-structure, we learned the first idea of the hierarchical brain from Hughlings Jackson and his thought that the relationship between different hierarchical regions is that higher areas inhibit lower areas. We showed the development of different kinds of the brain maps based cytoarchitectonic criteria. We also pointed out the existence of strong opposition opinions on the functional meanings based solely on cytoarchitectonic criteria or the detailed classification of brain maps. For the micro-structure, we showed the development of the classification of the different layers of the cortex. We pointed out that there are many ways to divide cortex into different layers. And since the neocortex is a biological tissue, there are no hard lines between different layers, but rather only different degrees of concentration. Furthermore, the computation based on the strict classification of different layers of the cortex is not very realistic since the neurons in one layer could also receive input in another layer.

For the connectivity of neocortex, we reviewed evidence for the hierarchical brain, the roles of feedforward and feedback connections and the convergence and divergence of the connections. For the hierarchical brain, we showed one of the foundations of modern neuroscience: the feedforward and feedback connections. On the roles of feedforward and feedback connections, we showed that the only type of neurons that can travel across different areas is the pyramidal neurons which suggested that for both feedforward and feedback connections, the projection neurons are always excitatory. For the target neurons, the electron microscope evidence showed that both feedforward and feedback connections mostly targeted on the excitatory neurons. The feedforward sends more input to the inhibitory

neurons (10% connections) than the feedback connections (less than 2% connections). These evidence suggested an excitatory loop in the feedforward and feedback connections. However, we also showed that even in the neurophysiological data, the inhibitory feedback effect has been found. For the convergence and divergence of the connections, we showed evidence of neuron tracing and axonal bifurcations suggesting that feedback is divergent.

For the temporal dynamic of neocortex, we reviewed the evidence on the time delay between areas and oscillations. We divided the time delay into two types: axonal conduction delay and response delay. We showed evidence that axonal conduction delay is very short (1-2ms) between different cortical areas with a very small jitter (less than 0.1ms). These properties allow an accurate and robust computation based on spike timing relationship between two areas. We also showed that the response delays between different areas are much longer (10-20ms for every stage of information processing). I proposed that the difference between the axonal conduction delay and response delay is the key computation window for the brain. For the oscillations, we showed the existence, the possible origins and the functional significance of oscillations (both power and phase).

At the end of the introduction of the previous neuroscience evidence, I combined the known information about the brain and tried to propose a canonical neural circuits.

From all of the evidence we reviewed, we could conclude the rules of the brain that we need to obey:

(1) Neurons and action potentials are the basics of every brain function. It is possible that the brain uses a complex method to implement these functions, however, the basics about the neurons could not be broken.
(2) The shapes of the neurons decided that only pyramidal neurons can carry information across areas and stellate neurons can only deal with local activities.
(3) The excitatory and inhibitory neurons are the basic functional elements of the brain.
(4) Feedforward and feedback connections are the basic connection types in the brain.
(5) Feedforward and feedback connections are mainly excitatory.
(6) Feedback connections are divergent.

We could also conclude some points that may be correct:

(1) The absence of layer 4 stellate neurons in non-sensory regions may suggest a weaker role for the stellate neurons for computation.
(2) It may be hard to learn the synaptic weights to the inhibitory neurons because of the absence of dendritic spines.
(3) We should rely less on the current classifications of the brain areas and cortical layers for their functional meanings. The laminar computation may be not particularly realistic.
(4) The excitatory recurrent feedforward and feedback network may be the typical structure in the brain.
(5) The difference between the axonal conduction delay and response delay may be the key computation window for the brain.

After the introduction of my understandings of the brain, I concluded my opinion about why I think predictive coding is a great model for the brain. I reviewed the development of the information theory's approach to understand the brain and suggested that predictive coding is a modern implementation of efficient coding theory and can be linked to the specific neural mechanism. I also reviewed the current literature about predictive coding, including the effects of predictive coding, its relationship with attention and oscillations. Taking into account these understandings of the brain, I proposed my motivation for my studies in my PhD.

In my PhD, I did three investigations: one theoretical investigation with the question "What is a better neuronal model for predictive coding under our current knowledge about the brain? " And two empirical investigations with the questions "What is the perceptual effect of predictive coding? " And "What is the relationship between oscillations and predictive coding? ".

In Chapter I, we showed the theoretical investigation of predictive coding, which is also the core of this thesis. Since we know that the classical predictive coding model does not constitute a neuronal model, we proposed a predictive coding model based on correlated spike times. The motivation for this study is from the interesting contradictions about feedback inhibition: feedback can have a selective and inhibitory effect, but feedback connections are divergent and excitatory. In this study, we demonstrated that it is possible to generate a selective inhibition effect by taking full advantage of the higher and lower area neurons' spike-time causality and phase/spike-time response curve, a fundamental neuronal response property.

In the simulations, we first showed that lower area neurons are less responsive to feedback excitation (relative inhibition) when their spike times are correlated with the active neurons in the higher area. The mechanism underlying it is based on different spike-time advances for different feedback time relative to lower neuron's last spike time. Predictable neurons (lower area neurons that are driving higher area neuron) receive feedback just after their last spike, thus, the feedback has very limited effect on their activity. On the other hand, the feedback time to unpredictable neurons (lower area neurons that are not driving higher area neuron) receives an average spike time advance. We then showed the four factors that can affect the feedback and therefore affect the proposed spike timing based selectivity: the feedback strength, the axonal conduction delay, the noise in the system and the predictability of the predictable neurons. We showed that feedback strengths modulate the selectivity in both ways, a monotonic relationship between the selectivity and axonal conduction delay (smaller the delay, bigger the effect), and between the selectivity and the predictability (more predictable lower area neurons create stronger selectivity). We also showed the strong resistance of such model to the noise in the system. Then, we showed that normalization in the lower area can turn the relative inhibition into absolute inhibition. The proposed computational principle provides a viable neuronal mechanism for efficient coding with a much more flexible spike-time based selectivity than traditional connection-weight based selectivity.

We then further asked the question about the role of the spike timing dependent plasticity in such model. We demonstrated that the spike-time relationship generated by the model can take advantage of the STDP to enhance the existing selective inhibitory effect.

In Chapter II, inspired by the excitatory feedback connections in the model, we employed a psychophysics approach for the perceptual effect of predictive coding since most studies using fMRI showed that predictive feedback is inhibitory.

To produce predictive feedback, we employed similar stimuli as in Murray et al.: 3D-shape outlines and random-lines versions of the same stimuli(Murray et al., 2002). The former can be readily recognized, and should thus normally produce more predictive feedback than the latter. The two kinds of stimuli (3D shape and random lines) were displayed on gray disks simultaneously on the left and right of a fixation point on a black background. Subjects were asked to compare the luminance of the two disks (report the side of the brightest disk). We obtained behavioral responses from 14 subjects (including 2 subjects with eye-tracker) and we found out a consistent behavioral response showing that the disk behind the 3D-shape stimulus was perceived brighter against the black background than the one behind the random-lines (meaningless) stimulus. Since previous evidence suggested a monotonic relationship between contrast perception and neural activity in early visual areas(Dean, 1981; Boynton et al., 1999), we interpret these results as evidence that predictive feedback had an excitatory effect on sensory activity as suggested in our model.

We performed control experiments to exclude three possible alternative explanations of our results: attention bias, local factors and response bias. Operations for the control experiments included replacing the center fixation point with an attentional demanding task (letter RSVP), reversing the polarity of the stimuli outline, from black to white, modifying the response instructions (asking “which disk was darker?” Instead of “which

disk was brighter?"), and changing the subjects' task (to a same/different perception task, by asking "Did the two disks have the same luminance?"). These control experiments showed that the alternative explanations of our results can be ruled out.

In Chapter III, we showed a study about the relationship between predictive coding and oscillations. Since predictive coding theory suggested an iterative nature of the interactions between lower- and higher-level areas, it is intuitive to assume that predictive coding also takes advantage of neural oscillations and predictions/prediction error could modulate sensory processing periodically. Since the phase could reflect the state of the oscillation, we investigated the relationship between the pre-stimulus phase (since it is not reset by the stimuli) and predictive coding's perceptual effect which we observed in the previous study.

We used a similar paradigm as the previous study to induce different amounts of predictive feedback (3D-shape and random-lines), and measured the corresponding effects on luminance judgment as trial-by-trial markers of the efficiency of predictive coding while the EEG activity was recorded. By analyzing the relationship between the post-stimulus decision and pre-stimulus EEG phase which is a reflection of the phase when prediction comes (after 3D-shape onset, the shape representation in higher area feeds back its prediction), we found that two pre-stimulus ongoing oscillations from different regions and frequencies could strongly influence the luminance judgment: a contralateral frontal (higher area) theta oscillation and a contralateral occipital (lower area) beta oscillation. The phase of the theta oscillation before stimulus onset could explain 14% of the luminance judgment difference while the

phase of the beta oscillation could explain 19%. Control analyses ruled out contamination of the phase-behavior relationship by post-stimulus activity or ocular artifacts. These results not only imply that predictive coding is a periodic process, but also reveal two periodicities with different sources: the brain sends back predictions in a theta frequency, and sends forward predictive errors at a beta frequency.

Strengths and weaknesses

In this thesis, I presented my current understanding about the limitations of the brain and three studies that are closely linked to predictive coding. I believe these studies have the following strengths and weaknesses:

Strengths

Neurophysiological facts driven research philosophy

In our path to find out the working principles of the brain, there are two main research philosophies:

The first one usually comes with the analogy such as “we can’t understand how the computer works by opening up the computer itself”. This analogy appears to be reasonable and is believed by many researchers in the traditional psychology research. It also implies that the neurophysiological evidence is not important for us to understand the brain and the only good way to understand it is to perform experiments

on waking subjects. Under this philosophy, the brain is treated as a black box and the researchers try to interpret its function by simply interacting with this black box. However, I believe that this is not an optimal way to understand the brain. It is possible to understand the working principle of a computer by looking into the detailed electrical designs, and this is extremely useful if we could combine the designs with information we obtained from the interaction of the computer. Without any knowledge from neuroscience, we will end up with a lot of very vague words about the brain, such as attention and consciousness, which no one really understands.

The second philosophy usually has the name of "connectome", in which the goal of such research is to obtain a comprehensive map of neural connections in the brain. Many researchers in neuroscience favor this philosophy and claim that the brain will be understood the day that we know every detail of the connections and the structure. Many detailed models of the connections (even in the resolution of the synapse level) were created. I believe that this is also not the optimal way to understand the brain. Even when we have the truthful and detailed data about the connections, since the data are huge and complicated, it is possible that we still cannot abstract the working principles from those data.

My philosophy of research is to combine the neurophysiological facts and empirical evidence with the guidance of computational theory such as efficient coding. The neurophysiological facts are the basic of the research since they are robust and stable. With the discovery of many hard working neurophysiologists, now it is hard to find out new and universal facts that we have not discovered yet. When there is a

contradiction between the neurophysiological evidence and the theory, we should not change the evidence to fit the theory, but rather the other way around. The combination of the neurophysiological facts and empirical evidence can help us to create a much better model for the brain.

Innovative theoretical model

In the theoretical model, we combined the causally related spike-time and the phase response curve, one fundamental property of the neuron to generate the selectivity. This selectivity also fit the prediction of the predictive coding theory.

The proposed model creatively solved the feedback selectivity problem. It is not difficult to imagine that STDP can help the feedforward pathway to generate the selectivity: if lower area neuron always drives the higher area neuron, their spike-timing relationship follows the requirement for an increasing weight. However, it is difficult to see the emergence of the feedback selectivity. In the proposed model, we used non-selective connections and generated the feedback selectivity using the causally related spike time.

The proposed model also solved the problem of robustness in temporal coding. The neural spike times are too variable to support robust computation: the exact spike-timing is random and the index of dispersion is considerable. The lack of robustness is the main reason why the neuroscience community believes in rate coding rather in temporal coding. In the proposed model, we used the spike-time correlation

rather than the absolute spike-time, which is much robust. The simulation showed that it can resist huge amounts of noise.

The proposed model can also fit well with the neurophysiological evidence. For example, the architecture of the model fit well with the evidence that feedback connections are excitatory and divergent. Furthermore, the phase/spike-time response curve is one of the fundamental properties of neurons. Thus, the computational mechanisms underlying the proposed spike-time based selectivity are well supported experimentally.

Excitatory and divergent feedback also fits the classical observation of "attention". The proposed model has a lot of similarity with the known features of "attention". The non-selective feedback fits the spot-light assumption of attention and the excitatory feedback fit the biased competition theory. Different higher area in the proposed model may correspond to different types of attention: if the higher area represents the low level features such as color, shape, the proposed model may correspond to feature-based attention, if the higher area represents the object, the proposed model may correspond to the object-based attention. The inhibitory effect of the model may link to the inhibition of return effect in attention. Thus, the proposed model may be also a viable mechanism for the classical observation of "attention".

Convincing empirical evidence

We obtained very convincing empirical evidence in the two experimental studies. For the study of the perceptual effect of predictive coding, we obtained consistent behavioral responses of 14 subjects and

we performed four control experiments to rule out the alternative explanations for our results. For the study of the relationship between the predictive coding and oscillation, we performed a non-parametric test of 80,000,000 simulations to obtain the significance of the observed effect. I consider these efforts made the observed empirical evidence convincing.

Weaknesses

Direct evidence is required

Even though we proposed an innovative and promising model of predictive coding, we still require more evidence to prove the proposed mechanism. It is hard to obtain the spike times relationship from two neurons in different areas since it requires simultaneous recordings in separate regions. Luckily, with improved recording methods and newly developed techniques such as optogenetics, now, it is possible to do these types of experiments. I believe that we can put the proposed computational model to the test in the near future.

Better experimental methods are required

In the psychophysics experiment, we found an opposite effect as the traditional predictive coding evidence: we showed that predictive feedback can have an excitatory role rather than inhibitory. The methods used in both our study and the previous study are not optimal: the previous fMRI method had a poor temporal resolution and even though the psychophysics have a good temporal resolution, it is not the most direct way to observe excitatory or inhibitory signals in the brain.

We hope a better experimental method can enable us to solve the problem using a direct and accurate manner. This problem is also applicable to the EEG study. Due to the nature of the EEG system, we can hardly measure the neural oscillations with a frequency higher than the beta band. We may conclude that this is the probable cause for the absence of gamma band oscillations in our analysis. I think that we can obtain a much more comprehensive profile of the relationship between oscillations and predictive coding if we can use a better way to measure oscillatory activity in the brain.

Perspective and future work

Rate coding vs. Temporal coding

Rate coding vs. Temporal coding is a long existing debate in the field of neuroscience. From 1920s, Edgar Adrian already observed that the firing rates of a frog muscle's stretch receptor increases as a function of the load on the muscle. Many experiments showed that the rate coding scheme is preferred since the spike trains for the same input stimulus usually have a similar firing rates in different trials. The rate coding scheme also is shown to have a functional role in perception (such as the direct link between the firing rate and the strength of the stimulus).

For temporal coding, the main arguments lie on three main points: (1) the response of the brain is too fast for rate coding. For example, experimental studies of neurons in various parts of the monkey brain

showed a selective response only 100-150 ms after stimulus onset, some neurons can have a selective response in only 80 – 90 ms. There is just not enough time for counting the number of spikes. (2) The very first spike for the stimuli jitter from trial to trial was less than 1 ms, thus, it is possible to use the first spike for a temporal coding. (3) Temporal coding obviously carries more information than rate coding. However, the fatal flaw of temporal coding is that the absolute spike time is random. Thus, it is basically impossible to use the spike time for any robust computation.

In our model, we proposed to use the relative spike time rather than the absolute spike time. This proposal increased the robustness of the temporal coding significantly. We considered that the only purpose of a single spike is to advance or delay the next spike time of the target neuron. The leaky nature of the membrane decided that this advancement is sensitive to the input time. Thus, by changing the temporal coding, we can modify the spiking rate, which is usually used as the indicator of neural activity. The proposed model redefines the concept of temporal coding and allows a much more robust and effective computation.

In the future, I think it worth to use the neurophysiological evidence to prove or disprove the proposed model. If the proposed model can be verified, this could be potentially a working principle of the brain.

Excitatory non-selective feedback vs. Inhibitory selective feedback

The debate between the excitatory non-selective feedback and inhibitory selective feedback relies on the researcher's philosophy to

deal with the relationship between the theory and neurophysiological evidence. It's true that there is evidence for a selective inhibitory feedback, not only in theory but also in fMRI experiments. However, the evidence at a fine resolution (such as studies in neuroanatomy and electron microscope evidence) should be the basic evidence we should follow. It is possible to have a non-selective excitatory physical connections and generate a selective inhibitory effect. However, it is impossible to change the neuroanatomical and electron microscope evidence simply because an opposite effect was found in the population-level. Should we change the evidence to fit the theory, or should we change the theory to fit the evidence? This seems to be a simple question, however, some scientists find it much easier to do the former one since it does not require explaining the contradictions between the existing model and evidence. I think one of the most interesting things in the research of neuroscience is to solve these contradictions, rather than avoid these contradictions. I think we can understand more about the brain using this method.

Of course, there are evidence showing the inhibitory feedback connections (some scientists are trying hard to find out evidence for that) or excitatory feedback targeted on lower area inhibitory neurons, however, the relative percentage should tell us more about the functional importance (e.g. much less inhibitory neurons are targeted by feedback connections than feedforward connections). For the non-selectivity, the non-selectivity is not only limited to the space domain (non-selective spatial effect), but also other domains (e.g. feature domain, object domain). In these conditions, the feedback from these areas should be able to produce non-selective feature or object effect.

In the future, I think that more evidence is required for us to understand the details and mechanism of the brain. I think it is interesting to use different methods to show that the neural activity in the population level follows the basics of neurophysiology using a specific neuronal mechanism. It is interesting to show both the evidence and propose better models that can fit both the neurophysiological evidence and the neural activity in the population level.

Attention vs. Expectation

Attention and expectation are two vague words from the field of psychology. Is it possible to distinguish the two words? Many researchers argued that expectation and attention belongs to two distinct brain operations. However, when they perform the experiments, they usually used very similar tasks to achieve the expectation and attention operations. Sometimes, the same operation was used in two papers while one paper calls it expectation and the other calls it attention. Furthermore, it is worth noticing that there are different kinds of attention also (e.g. feature-based attention, object-based attention, etc.).

Firstly, I think it is useful to avoid to use the words such as “attention” and “expectation” since they describe a subjective feeling rather than an objective measure. I think it is better to call them “feedback” since we can know clearly that it means the signal sending from higher area to lower area. Secondly, I think it is possible and better to combine all the expectation and attention claims into a more explicit manner. For example, it would be considerably better to label every expectation operation and attention operation by their features, such as the type of the cues, the appearance time of the cues and so on. At last, we should

acknowledge that the feedback may come from different higher areas and have different features. We can describe the possible origins (such as a moving face cue may suggest the feedback may come from the MT and FFA) in the higher area and study the effects of these different types of feedback.

In the future, we can investigate more on the relationship between predictive coding and attention. As suggested above, with the new name feedback, we could investigate the effect of different types of feedback (e.g. different modalities, different feature levels' feedback).

Conclusion

In this thesis, I presented my research on predictive coding in my PhD. I tried to address one of the key problems in predictive coding: predictive feedback. With the guidance of evidence from neurophysiology, I proposed the nature of feedback (excitatory) and the general modulation characters (non-selectivity). We proposed a creative model to implement predictive coding using the phase response curve and causally related spike times between the higher area neurons and predictable neurons in the lower area. We also showed that the classical STDP rule can enhance the selectivity created by the spike times. Two empirical evidence was also showed in the thesis to discuss the relationship among predictive coding, perception and oscillations.

Robust system with complicated functions usually follows simple basic rules. Any system with complex rules are vulnerable to the unstable environment. Human brains are working in more than 7 billion bodies with very low defect rate. Thus, it is reasonable to assume simple rules for the human brain. These simple rules should not be determined by the subjective experience (e.g. attention, consciousness), but rather the objective observations (e.g. neuron, feedforward and feedback). The proposed model takes advantages of one fundamental property of neurons in the temporal domain (phase response curve) and the properties of feedback connections (non-selective excitatory). The causally related spike-times created the selective inhibitory effect in predictive coding, a modern theory of efficient coding. The proposed model has its significance in the process of understanding the brain. The final answers of the brain should not be and will not be too complicated for us to understand.

Reference

Abbott LF, Nelson SB (2000) Synaptic plasticity: taming the beast. Nat Neurosci 3 Suppl:1178–1183.

Abeles M (1991) Corticonics: Neural Circuits of the Cerebral Cortex.

Adrian ED (1926) The impulses produced by sensory nerve endings: Part I. J Physiol 61:49–72.

Albrecht DG, Hamilton DB (1982) Striate cortex of monkey and cat: contrast response function. J Neurophysiol 48:217–237.

Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L (2010) Stimulus Predictability Reduces Responses in Primary Visual Cortex. J Neurosci 30:2960–2966.

Allman J, Miezin F, McGuinness E (1985) Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. Annu Rev Neurosci 8:407–430.

Anderson JS, Carandini M, Ferster D (2000) Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. J Neurophysiol 84:909–926.

Angelucci A, Bullier J (2003) Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? J Physiol Paris 97:141–154.

Angelucci A, Levitt JB, Walton EJS, Hupe J-M, Bullier J, Lund JS (2002) Circuits for local and global signal integration in primary visual cortex. J Neurosci 22:8633–8646.

Arnal LH, Giraud AL (2012) Cortical oscillations and sensory predictions. Trends Cogn Sci 16:390–398.

Arnal LH, Wyart V, Giraud A-L (2011) Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nat Publ Gr 14:797–801.

Atallah B V, Scanziani M (2009) Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. Neuron 62:566–577.

Attneave F (1954) Some informational aspects of visual perception. Psychol Rev 61:183–193.

Azouz R, Gray CM (2000) Dynamic spike threshold reveals a mechanism for synaptic coincidence detection in cortical neurons in vivo. Proc Natl Acad Sci U S A 97:8110–8115.

Bailey P, Bonin G von (1951) The isocortex of man.

Barbas H, Rempel-Clower N (1997) Cortical structure predicts the pattern of corticocortical connections. Cereb Cortex 7:635–646.

Barlow HB (1961a) Possible principles underlying the transformation of sensory messages. Sens Commun:217–234.

Barlow HB (1961b) Possible principles underlying the transformation of sensory messages. Sensory communication (pp. 217--234).

Barlow HB (1972) Single units and sensation: a neuron doctrine for perceptual psychology? Perception 1:371–394.

Barone P, Batardiere a, Knoblauch K, Kennedy H (2000) Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. J Neurosci 20:3263–3281.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. Neuron 76:695–711.

Bauer M, Stenner MP, Friston KJ, Dolan RJ (2014) Attentional Modulation of Alpha/Beta and Gamma Oscillations Reflect Functionally Distinct Processes. J Neurosci 34:16117–16125.

Bean BP (2007) The action potential in mammalian central neurons. Nat Rev Neurosci 8:451–465.

Bekisz M, Wróbel A (2003) Attention-dependent coupling between beta activities recorded in the cat's thalamic and cortical representations of the central visual field. Eur J Neurosci 17:421–426.

Benes FM, Berretta S (2001) GABAergic interneurons: implications for understanding schizophrenia and bipolar disorder. Neuropsychopharmacology 25:1–27.

Bi G, Poo MM (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci 18:10464–10472.

Billings-Gagliardi S, Chan-Palay V, Palay SL (1974) A review of lamination in area 17 of the visual cortex of Macaca mulatta. J Neurocytol 3:619–629.

Bishop G (1932) Cyclic changes in excitability of the optic pathway of the rabbit. Am J Physiol Content.

Blakemore C, Tobin E (1972) Lateral inhibition between orientation detectors in the cat's visual cortex. Exp Brain Res 15.

Bohte SM (2004) The evidence for neural information processing withprecise spike-times: A survey. Nat Comput 3:195–206.

Bonin G von (1942) The striate area of primates. J Comp Neurol.

Bonin G von, Bailey P (1947) The Neocortex of Macaca Mulatta.

Bosman C a., Schoffelen JM, Brunet N, Oostenveld R, Bastos AM, Womelsdorf T, Rubehn B, Stieglitz T, De Weerd P, Fries P (2012) Attentional Stimulus Selection through Selective Synchronization between Monkey Visual Areas. Neuron 75:875–888.

Boynton GM, Demb JB, Glover GH, Heeger DJ (1999) Neuronal basis of contrast discrimination. Vision Res 39:257–269.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Braun J (1998) Vision and attention: the role of training. Nature 393:424–425.

Brette R, Gerstner W (2005) Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. J Neurophysiol 94:3637–3642.

Britten K, Shadlen M (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. J ….

Britten K, Shadlen M (1993) Responses of neurons in macaque MT to stochastic motion signals. Vis ….

Brodmann K (1905) Beiträge zur histologischen Lokalisation der Grosshirnrinde. Vierte Mitteilung: die Riesenpyramidentypus und sein Verhalten zu den Furchen bei den. J Psychol Neurol.

Brodmann K (1909) Localisation in the Cerebral Cortex.

Buffalo E a, Fries P, Landman R, Buschman TJ, Desimone R (2011) Laminar differences in gamma and alpha coherence in the ventral stream. Proc Natl Acad Sci U S A 108:11262–11267.

Buracas G, Zador A, DeWeese M, Albright T (1998) Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. Neuron.

Busch N a, Dubois J, VanRullen R (2009) The phase of ongoing EEG oscillations predicts visual perception. J Neurosci 29:7869–7876.

Busch NA, VanRullen R (2010) Spontaneous EEG oscillations reveal periodic sampling of visual attention. Proc Natl Acad Sci U S A 107:16048–16053.

Buschman TJ, Miller EK (2007) Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. Science (80-) 315:1860–1862.

Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. Science 304:1926–1929.

Buzsáki G, Wang X-J (2012) Mechanisms of Gamma Oscillations. Annu Rev Neurosci 35:203–225.

Cajal S y (1888) Estructura de los centros nerviosos de las aves.

Cajal S y (1899) Comparative study of the sensory areas of the human cortex.

Campbell AW (1905) Histological Studies on the Localisation of Cerebral Function.

Caporale N, Dan Y (2008) Spike timing-dependent plasticity: a Hebbian learning rule. Annu Rev Neurosci 31:25–46.

Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. Nat Rev Neurosci.

Carrasco MM, Carrasco MM, Ling S, Ling S, Read S, Read S (2004) Attention alters appearance. Nat Neurosci 7:308–313.

Churchill JD, Tharp JA, Wellman CL, Sengelaub DR, Garraghty PE (2004) Morphological correlates of injury-induced reorganization in primate somatosensory cortex. BMC Neurosci 5:43.

Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36:181–204.

Clements JD (1996) Transmitter timecourse in the synaptic cleft: Its role in central synaptic function. Trends Neurosci 19:163–171.

Clements JD, Lester R a, Tong G, Jahr CE, Westbrook GL (1992) The time course of glutamate in the synaptic cleft. Science (80-) 258:1498–1501.

Coogan TA, Burkhalter A (1993) Hierarchical organization of areas in rat visual cortex. J Neurosci 13:3749–3772.

Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE (1991) Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. J Neurosci 11:2383–2402.

Corbetta M, Shulman GL, Miezin FM, Petersen SE (1995) Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. Science (80-) 270:802–805.

De Carlos J a., Borrell J (2007) A historical reflection of the contributions of Cajal and Golgi to the foundations of neuroscience. Brain Res Rev 55:8–16.

Dean AF (1981) The relationship between response amplitude and contrast for cat striate cortical neurones. J Physiol 318:413–427.

DeAngelis GC, Freeman RD, Ohzawa I (1994) Length and width tuning of neurons in the cat's primary visual cortex. J Neurophysiol 71:347–374.

Delorme A, Makeig S (2004) EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21.

Deschenes M (1977) Dual origin of fibers projecting from motor cortex to SI in cat. Brain Res 132:159–162.

Douglas RJ, Martin K a C (2007) Mapping the matrix: the ways of neocortex. Neuron 56:226–238.

Douglas RJ, Martin KA (1991) A functional microcircuit for cat visual cortex. J Physiol 440:735–769.

Douglas RJ, Martin KAC (2004) Neuronal circuits of the neocortex. Annu Rev Neurosci 27:419–451.

Douglas RJ, Martin KAC, Whitteridge D (1989) A canonical microcircuit for neocortex. Neural Comput 1:480–488.

Drewes J, VanRullen R (2011) This Is the Rhythm of Your Eyes: The Phase of Ongoing Electroencephalogram Oscillations Modulates Saccadic Reaction Time. J Neurosci 31:4698–4708.

Dugué L, Marque P, VanRullen R (2011) The Phase of Ongoing Oscillations Mediates the Causal Relation between Brain Excitation and Visual Perception. J Neurosci 31:11889–11893.

Dugué L, Marque P, VanRullen R (2015) Theta Oscillations Modulate Attentional Search Performance Periodically. J Cogn Neurosci:945–958.

Economo C von, Koskinas G (1925) Die cytoarchitektonik der hirnrinde des erwachsenen menschen.

Egner T, Monti JM, Summerfield C (2010) Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream. J Neurosci 30:16601–16608.

Ermentrout B (1996) Type I membranes, phase resetting curves, and synchrony. Neural Comput 8:979–1001.

Fanselow EE, Richardson K a, Connors BW (2008) Selective, state-dependent activation of somatostatin-expressing inhibitory interneurons in mouse neocortex. J Neurophysiol 100:2640–2652.

Fatterpekar GM, Naidich TP, Delman BN, Aguinaldo JG, Gultekin SH, Sherwood CC, Hof PR, Drayer BP, Fayad Z a. (2002) Cytoarchitecture of the human cerebral cortex: MR microscopy of excised specimens at 9.4 Tesla. Am J Neuroradiol 23:1313–1321.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47.

Ferraina S, Paré M, Wurtz RH (2002) Comparison of cortico-cortical and cortico-collicular signals for the generation of saccadic eye movements. J Neurophysiol 87:845–858.

Ferrer JMR, Price DJ, Blakemore C (1988) The organization of corticocortical projections from area 17 to area 18 of the cat's visual cortex. Proc R Soc London B Biol Sci 233:77–98.

Fiebelkorn IC, Saalmann YB, Kastner S (2013) Rhythmic sampling within and between objects despite sustained attention at a cued location. Curr Biol 23:2553–2558.

Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am A 4:2379–2394.

Fischl B, Dale a M (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci U S A 97:11050–11055.

Flechsig P (1920) Anatomie des menschlichen Gehirns und Rückenmarks auf myelogenetischer Grundlage.

Fontaine B, Peña JL, Brette R (2014) Spike-Threshold Adaptation Predicted by Membrane Potential Dynamics In Vivo. PLoS Comput Biol 10:1–11.

Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud A-L (2014) The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nat Commun 5:4694.

Friedman DP (1983) Laminar patterns of termination of cortico-cortical afferents in the somatosensory system. Brain Res 273:147–151.

Fries P, Nikolić D, Singer W (2007) The gamma cycle. Trends Neurosci 30:309–316.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815–836.

Friston K, Kiebel S (2009) Predictive coding under the free-energy principle. Philos Trans R Soc Lond B Biol Sci 364:1211–1221.

Galán R, Ermentrout G, Urban N (2005a) Efficient Estimation of Phase-Resetting Curves in Real Neurons and its Significance for Neural-Network Modeling. Phys Rev Lett 94:158101.

Galán RF, Ermentrout GB, Urban NN (2005b) Efficient estimation of phase-resetting curves in real neurons and its significance for neural-network modeling. Phys Rev Lett 94:1–4.

Gandhi SP, Heeger DJ, Boynton GM (1999) Spatial attention affects brain activity in human primary visual cortex. Proc Natl Acad Sci 96:3314–3319.

Garey L (1971) A light and electron microscopic study of the visual cortex of the cat and monkey. Proc R Soc ....

Gilbert CD, Wiesel TN (1990) The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. Vision Res 30:1689–1701.

Girard P, Bullier J (1989) Visual activity in area V2 during reversible inactivation of area 17 in the macaque monkey. J Neurophysiol 62:1287–1302.

Girard P, Hupé JM, Bullier J (2001) Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. J Neurophysiol 85:1328–1331.

Girard P, Salin PA, Bullier J (1991a) Visual activity in areas V3a and V3 during reversible inactivation of area V1 in the macaque monkey. J Neurophysiol 66:1493–1503.

Girard P, Salin PA, Bullier J (1991b) Visual activity in macaque area V4 depends on area 17 input. Neuroreport 2:81–84.

Girard P, Salin PA, Bullier J (1992) Response selectivity of neurons in area MT of the macaque monkey during reversible inactivation of area V1. J Neurophysiol 67:1437–1446.

Glickstein M (1988) The discovery of the visual cortex. Sci Am 259:118–127.

Goldberg J a, Deister C a, Wilson CJ (2007) Response properties and synchronization of rhythmically firing dendritic neurons. J Neurophysiol 97:208–219.

Goodyear BG, Menon RS (1998) Effect of luminance contrast on BOLD fMRI response in human primary visual areas. J Neurophysiol 79:2204–2207.

Grothe I, Neitzel SD, Mandon S, Kreiter AK (2012) Switching neuronal inputs by differential modulations of gamma-band phase-coherence. J Neurosci 32:16172–16180.

Guillery RW, Sherman Sm (2001) Thalamic Relay Functions and Their Role in Corticocortical Communication:Generalizations from the Visual System. Neuron 33:163–175.

Haider B, Duque A, Hasenstaub AR, McCormick DA (2006) Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. J Neurosci 26:4535–4545.

Han B, VanRullen R (2014) Predictive Coding of Shape Affects the Perceived Luminance of the Surrounding Region. J Vis 14:72–72.

Hansel D, Mato G, Meunier C (1995) Synchrony in excitatory neural networks. Neural Comput 7:307–337.

Hanslmayr S, Aslan A, Staudigl T, Klimesch W, Herrmann CS, Bäuml K-H (2007) Prestimulus oscillations predict visual perception performance between and within subjects. Neuroimage 37:1465–1473.

Harris KD, Shepherd GMG (2015) The neocortical circuit: themes and variations. Nat Neurosci 18:170–181.

Harrison LM, Stephan KE, Rees G, Friston KJ (2007) Extra-classical receptive field effects measured in striate cortex with fMRI. Neuroimage 34:1199–1208.

Hebb DO (2005) The Organization of Behavior: A Neuropsychological Theory. Psychology Press.

Henry GH, Salin P a., Bullier J (1991) Projections from Areas 18 and 19 to Cat Striate Cortex: Divergence and Laminar Specificity. Eur J Neurosci 3:186–200.

Henze D a., Buzsáki G (2001) Action potential threshold of hippocampal pyramidal cells in vivo is increased by recent spiking activity. Neuroscience 105:121–130.

Hering H, Sheng M, Medical HH (2001) Dentritic spines: structure, dynamics and regulation. Nat Rev Neurosci 2:880–888.

Hodgkin A, Huxley a. F (1952) Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo. J Physiol 116:449–472.

Howard MA, Rubel EW (2010) Dynamic spike thresholds during synaptic integration preserve and enhance temporal response properties in the avian cochlear nucleus. J Neurosci 30:12063–12074.

Huang Y, Chen L, Luo H (2015) Behavioral Oscillation in Priming: Competing Perceptual Predictions Conveyed in Alternating Theta-Band Rhythms. J Neurosci 35:2830–2837.

Huang Y, Rao RPN (2011) Predictive coding. Wiley Interdiscip Rev Cogn Sci 2:580–593.

Hubel D, Wiesel T (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol:106–154.

Hubel D, Wiesel T (1965) Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J Neurophysiol 28:229–289.

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol 195:215–243.

Hupe J-M, Hupe J-M, James AC, James AC, Girard P, Girard P, Lomber SG, Lomber SG, Payne BR, Payne BR, Bullier J, Bullier J (2001) Feedback connections act on the early part of the responses in monkey visual cortex. J Neurophysiol 85:134–145.

Hupé JM, James a C, Payne BR, Lomber SG, Girard P, Bullier J (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature 394:784–787.

Isaacson JS, Isaacson JS, Scanziani M, Scanziani M (2011) How inhibition shapes cortical activity. Neuron 72:231–243.

Jackson J (1882) On some implications of dissolution of the nervous system. Med Press Circ.

Jellison BJ, Field AS, Medow J, Lazar M, Salamat MS, Alexander AL (2004) Diffusion Tensor Imaging of Cerebral White Matter: A Pictorial Review of Physics, Fiber Tract Anatomy, and Tumor Imaging Patterns. Am J Neuroradiol 25:356–369.

Jensen O, Bonnefond M, VanRullen R (2012) An oscillatory mechanism for prioritizing salient unattended stimuli. Trends Cogn Sci 16:200–206.

Jiang J, Summerfield C, Egner T (2013) Attention Sharpens the Distinction between Expected and Unexpected Percepts in the Visual Brain. J Neurosci 33:18438–18447.

Johnson RR, Burkhalter A (1996) Microcircuitry of forward and feedback connections within rat visual cortex. J Comp Neurol 368:383–398.

Johnson RR, Burkhalter A (1997) A polysynaptic feedback circuit in rat visual cortex. J Neurosci 17:7129–7140.

Joseph JS, Chun MM, Nakayama K (1997) Attentional requirements in a “preattentive” feature search task. Nature 387:805–807.

Joshi S (2007) Structure-Function Relations in Macaque V1.

Kätzel D, Zemelman B V, Buetfering C, Wölfel M, Miesenböck G (2011) The columnar and laminar organization of inhibitory connections to neocortical excitatory cells. Nat Neurosci 14:100–107.

Kendal ER, Schwartz JH, Jessell TM, R Kandel E, Harris Schwartz J, M Jessell T (2000) Principles of neural science. :1414.

Kennedy H, Bullier J (1985) A double-labeling investigation of the afferent connectivity to cortical areas V1 and V2 of the macaque monkey. J Neurosci 5:2815–2830.

Kisvarday ZF, Cowey A, Somogyi P (1986) Synaptic relationships of a type of GABA-immunoreactive neuron (clutch cell), spiny stellate cells and lateral geniculate nucleus afferents in layer IVC of the monkey striate cortex. Neuroscience 19:741–761.

Knierim JJ, van Essen DC (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J Neurophysiol 67:961–980.

Koch C (1998) Biophysics of computation: information processing in single neurons. Oxford university press.

Koch C, Poggio T (1999) Predicting the visual world: silence is golden. Nat Neurosci 2:9–10.

Kok P, de Lange FP (2014) Shape Perception Simultaneously Up- and Downregulates Neural Activity in the Primary Visual Cortex. Curr Biol 24:1531–1535.

Kok P, Jehee JFM, de Lange FP (2012a) Less is more: expectation sharpens representations in the primary visual cortex. Neuron 75:265–270.

Kok P, Kok P, Rahnev D, Rahnev D, Jehee JFM, Jehee JFM, Lau HC, Lau HC, de Lange FP, de Lange FP (2012b) Attention reverses the effect of prediction in silencing sensory signals. Cereb Cortex 22:2197–2206.

Kötter R, Wanke E (2005) Mapping brains without coordinates. Philos Trans R Soc Lond B Biol Sci 360:751–766.

Kwag J, Paulsen O (2009) The timing of external input controls the sign of plasticity at local synapses. Nat Neurosci 12:1219–1221.

Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. Hum Brain Mapp 208:194–208.

Landau AN, Fries P (2012) Attention samples stimuli rhythmically. Curr Biol 22:1000–1004.

Larkman a, Mason a (1990) Correlations between morphology and electrophysiology of pyramidal neurons in slices of rat visual cortex. I. Establishment of cell classes. J Neurosci 10:1407–1414.

Larkum M (2013) A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. Trends Neurosci 36:141–151.

Lengyel M, Kwag J, Paulsen O, Dayan P (2005) Matching storage and recall: hippocampal spike timing-dependent plasticity and phase response curves. Nat Neurosci 8:1677–1683.

Levitt JB, Lund JS (1997) Contrast dependence of contextual effects in primate visual cortex. Nature 387:73–76.

Lewis J, Essen D Van (2000) Mapping of architectonic subdivisions in the macaque monkey, with emphasis on parieto - occipital cortex. J Comp Neurol.

Liu Y-J, Ehrengruber MU, Negwer M, Shao H-J, Cetin AH, Lyon DC (2013) Tracing inputs to inhibitory or excitatory neurons of mouse and cat visual cortex with a targeted rabies virus. Curr Biol 23:1746–1755.

Lopes da Silva FH, van Rotterdam a, Storm van Leeuwen W, Tielen a M (1970) Dynamic characteristics of visual evoked potentials in the dog. II. Beta frequency selectivity in evoked potentials and background activity. Electroencephalogr Clin Neurophysiol 29:260–268.

López-Muñoz F, Boya J, Alamo C (2006) Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel Prize award to Santiago Ramón y Cajal. Brain Res Bull 70:391–405.

Lui JH, Hansen D V., Kriegstein AR (2011) Development and evolution of the human neocortex. Cell 146:18–36.

Maier A, Adams GK, Aura C, Leopold D a (2010) Distinct superficial and deep laminar domains of activity in the visual cortex during rest and stimulation. Front Syst Neurosci 4:1–11.

Manzoni T, Caminiti R, Spidalieri G, Morelli E (1979) Brain Anatomical and Functional Aspects of the Associative Projections from Somatic Area SI to SII *. Exp Brain Res 470:453–470.

Markov NT et al. (2014) A weighted and directed interareal connectivity matrix for macaque cerebral cortex. Cereb Cortex 24:17–36.

Markov NT, Ercsey-Ravasz M, Van Essen DC, Knoblauch K, Toroczkai Z, Kennedy H (2013a) Cortical High-Density Counterstream Architectures. Science (80-) 342:1238406.

Markov NT, Misery P, Falchier a., Lamy C, Vezoli J, Quilodran R, Gariel M a., Giroud P, Ercsey-Ravasz M, Pilaz LJ, Huissoud C, Barone P, Dehay C, Toroczkai Z, Van Essen DC, Kennedy H, Knoblauch K (2011) Weight consistency specifies regularities of macaque cortical networks. Cereb Cortex 21:1254–1272.

Markov NT, Vezoli J, Chameau P, Falchier A, Quilodran R, Huissoud C, Lamy C, Misery P, Giroud P, Ullman S, Barone P, Dehay C, Knoblauch K, Kennedy H (2013b) Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. J Comp Neurol 522:225–259.

Markram H, Markram H, Toledo-Rodriguez M, Toledo-Rodriguez M, Wang Y, Wang Y, Gupta A, Gupta A, Silberberg G, Silberberg G, Wu C, Wu C (2004) Interneurons of the neocortical inhibitory system. Nat Rev Neurosci 5:793–807.

Markram H, Tsodyks M (1996) Redistribution of synaptic efficacy between neocortical pyramidal neurons. Nature 382:807–810.

Mason a, Larkman a (1990) Correlations between morphology and electrophysiology of pyramidal neurons in slices of rat visual cortex. II. Electrophysiology. J Neurosci 10:1415–1428.

Mason a, Nicoll a, Stratford K (1991) Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro. J Neurosci 11:72–84.

McAdams CJ, Maunsell JHR (1999) Effects of Attention on the Reliability of Individual Neurons in Monkey Visual Cortex. Neuron 23:765–773.

Mountcastle V (1995) The evolution of ideas concerning the function of the neocortex. Cereb Cortex 5:289–295.

Movshon JA, Movshon JA, Newsome WT, Newsome WT (1996) Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J Neurosci 16:7733–7741.

Mumford D, Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biol Cybern 66:241–251.

Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL (2002) Shape perception reduces activity in human primary visual cortex. Proc Natl Acad Sci U S A 99:15164–15169.

Nassi JJ, Lomber SG, Born RT (2013) Corticocortical feedback contributes to surround suppression in V1 of the alert primate. J Neurosci 33:8504–8517.

Nelson JI, Frost BJ (1978) Orientation-selective inhibition from beyond the classic visual receptive field. Brain Res 139:359–365.

Netoff TI, Banks MI, Dorval AD, Acker CD, Haas JS, Kopell N, White J a (2004) Synchronization in hybrid neuronal networks of the hippocampal formation. J Neurophysiol 93:1197–1208.

Nowak L, Bullier J (1997) The Timing of Information Transfer in the Visual System. In: Extrastriate Cortex in Primates SE - 5 (Rockland K, Kaas J, Peters A, eds), pp 205–241 Cerebral Cortex. Springer US.

Nowak LG, James AC, Bullier J (1997) Corticocortical connections between visual areas 17 and 18a of the rat studied in vitro: spatial and temporal organisation of functional synaptic responses. Exp Brain Res 117:219–241.

Okun M, Lampl I (2008) Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. Nat Neurosci 11:535–537.

Olsen SR, Bortone DS, Adesnik H, Scanziani M (2012) Gain control by layer six in cortical circuits of vision. Nature 483:47–52.

Olshausen BA (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609.

Passingham RE, Wise SP (2012) The Neurobiology of the Prefrontal Cortex: Anatomy, Evolution, and the Origin of Insight. OUP Oxford.

Paxinos G, Huang X, Toga A (1999) The rhesus monkey brain in stereotaxic coordinates.

Peters A, Kara DA (1985a) The neuronal composition of area 17 of rat visual cortex. I. The pyramidal cells. J Comp Neurol 234:218–241.

Peters A, Kara DA (1985b) The neuronal composition of area 17 of rat visual cortex. II. The nonpyramidal cells. J Comp Neurol 234:242–263.

Petersen CC, Petersen CC, Sakmann B, Sakmann B (2000) The excitatory neuronal network of rat layer 4 barrel cortex. J Neurosci 20:7579–7586.

Phillips JM, Vinck M, Everling S, Womelsdorf T (2014) A long-range fronto-parietal 5- to 10-Hz network predicts “top-down” controlled guidance in a task-switch paradigm. Cereb Cortex 24:1996–2008.

Poo C, Isaacson JS (2009) Odor representations in olfactory cortex: “sparse” coding, global inhibition, and oscillations. Neuron 62:850–861.

Preyer AJ, Butera RJ (2005) Neuronal oscillators in aplysia californica that demonstrate weak coupling in vitro. Phys Rev Lett 95:1–4.

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.

Reddy L, Kanwisher NG, VanRullen R (2009) Attention and biased competition in multi-voxel object representations. Proc Natl Acad Sci U S A 106:21447–21452.

Reyes AD, Fetz EE (1993) Effects of transient depolarizing potentials on the firing rate of cat neocortical neurons. J Neurophysiol 69:1673–1683.

Rockland KS, Pandya DN (1979) Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. Brain Res 179:3–20.

Rodman HR, Gross CG, Albright TD (1989) Afferent basis of visual response properties in area MT of the macaque. I. Effects of striate cortex removal. J Neurosci 9:2033–2050.

Rohenkohl G, Cravo AM, Wyart V, Nobre AC (2012) Temporal Expectation Improves the Quality of Sensory Information. J Neurosci 32:8424–8428.

Roopun AK, Middleton SJ, Cunningham MO, LeBeau FEN, Bibbig A, Whittington M a, Traub RD (2006) A beta2-frequency (20-30 Hz) oscillation in nonsynaptic networks of somatosensory cortex. Proc Natl Acad Sci U S A 103:15646–15650.

Roy SA, Alloway KD (2001) Coincidence Detection or Temporal Integration? What the Neurons in Somatosensory Cortex Are Doing. J Neurosci 21:2462–2473.

Salin P, Bullier J (1995) Corticocortical connections in the visual system: structure and function. Physiol Rev.

Sandell JH, Schiller PH (1982) Effect of cooling area 18 on striate cortex cells in the squirrel monkey. J Neurophysiol 48:38–48.

Sani I, Santandrea E, Golzar a., Morrone MC, Chelazzi L (2013) Selective Tuning for Contrast in Macaque Area V4. J Neurosci 33:18583–18596.

Schneider DM, Nelson A, Mooney R (2014) A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature 513:189–194.

Schultheiss NW, Prinz AA, Butera RJ (2012) Phase Response Curves in Neuroscience. Springer Science & Business Media.

Shadlen M, Newsome W (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J Neurosci.

Shao Z, Burkhalter A (1996) Different balance of excitation and inhibition in forward and feedback circuits of rat visual cortex. J Neurosci 16:7353–7365.

Shepherd GM (2004) The synaptic organization of the brain. Sourcebooks, Inc.

Sholl DA (1956) The organization of the cerebral cortex.

Silberberg G, Markram H (2007) Disynaptic Inhibition between Neocortical Pyramidal Cells Mediated by Martinotti Cells. Neuron 53:735–746.

Smeal RM, Ermentrout GB, White JA (2010) Phase-response curves and synchronized neural networks. Philos Trans R Soc Lond B Biol Sci 365:2407–2422.

Softky W, Koch C (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. J Neurosci 13:334–350.

Song K, Meng M, Chen L, Zhou K, Luo H (2014) Behavioral Oscillations in Attention: Rhythmic Alpha Pulses Mediated through Theta Band. J Neurosci 34:4837–4844.

Song S, Miller KD, Abbott LF (2000) Competitive Hebbian learning through spike-time-dependent synaptic plasticity. Nat Neurosci 3:919–926.

Spratling MW (2008a) Predictive coding as a model of biased competition in visual attention. Vision Res 48:1391–1408.

Spratling MW (2008b) Reconciling Predictive Coding and Biased Competition Models of Cortical Function. Front Comput Neurosci 2.

Spruston N (2009) Pyramidal neuron. Scholarpedia 4:6130.

Summerfield C, de Lange FP (2014) Expectation in perceptual decision making: neural and computational mechanisms. Nat Rev Neurosci.

Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. Trends Cogn Sci 13:403–409.

Summerfield C, Egner T, Greene M, Koechlin E, Mangels J, Hirsch J (2006) Predictive codes for forthcoming perception in the frontal cortex. Science (80-) 314:1311–1314.

Summerfield C, Trittschuh EH, Monti JM, Mesulam M-M, Egner T (2008) Neural repetition suppression reflects fulfilled perceptual expectations. Nat Publ Gr 11:1004–1006.

Summerfield C, Wyart V, Johnen VM, de Gardelle V (2011) Human Scalp Electroencephalography Reveals that Repetition Suppression Varies with Expectation. Front Hum Neurosci 5:67.

Susan Standring, PhD Ds (2009) Gray's Anatomy 40th edition.

Swadlow HA, Weyand TG (1981) Efferent systems of the rabbit visual cortex: Laminar distribution of the cells of origin, axonal conduction velocities, and identification of axonal branches. J Comp Neurol 203:799–822.

Thorpe SJ, Fabre-Thorpe M (2001) Seeking categories in the brain. Sci (New York, NY):260–262.

Tiesinga P, Sejnowski TJ (2009) Cortical Enlightenment: Are Attentional Gamma Oscillations Driven by ING or PING? Neuron 63:727–732.

Todorovic A, van Ede F, Maris E, de Lange FP (2011) Prior Expectation Mediates Neural Adaptation to Repeated Sounds in the Auditory Cortex: An MEG Study. J Neurosci 31:9118–9123.

Tolhurst D (1989) The amount of information transmitted about contrast by neurones in the cat's visual cortex. Vis Neurosci.

Tolhurst D, Movshon J, Dean A (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Res.

Tovee M, Rolls E (1993) Information encoding and the responses of single neurons in the primate temporal visual cortex. J ….

Toyama K, Matsunami K, Ono T, Tokashiki S (1974) An intracellular study of neuronal organization in the visual cortex. Exp brain Res 21:45–66.

Tripathy SJ, Burton SD, Geramita M, Gerkin RC, Urban NN (2015) Brain-wide analysis of electrophysiological diversity yields novel categorization of mammalian neuron types. J Neurophysiol 113:3474–3489.

Tsubo Y, Takada M, Reyes AD, Fukai T (2007) Layer and frequency dependencies of phase response properties of pyramidal neurons in rat motor cortex. Eur J Neurosci 25:3429–3441.

Van Essen D, Maunsell J (1980) Two-dimensional maps of the cerebral cortex. J Comp Neurol 191:255–281.

Van Essen DC, Maunsell JHR (1983) Hierarchical organization and functional streams in the visual cortex. Trends Neurosci 6:370–375.

van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis M-A, Poort J, van der Togt C, Roelfsema PR (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc Natl Acad Sci U S A 111:14332–14341.

Van Rullen R, Thorpe SJ (2001) Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. Neural Comput 13:1255–1283.

VanRullen R (2013) Visual attention: a rhythmic process? Curr Biol 23:R1110–R1112.

VanRullen R, Busch NA, Drewes J, Dubois J (2011) Ongoing EEG Phase as a Trial-by-Trial Predictor of Perceptual and Attentional Variability. Front Psychol 2:60.

VanRullen R, Carlson T, Cavanagh P (2007) The blinking spotlight of attention. Proc Natl Acad Sci U S A 104:19204–19209.

VanRullen R, Thorpe SJ (2002) Surfing a spike wave down the ventral stream. Vision Res 42:2593–2615.

VanRullen R, VanRullen R, Guyonneau R, Guyonneau R, Thorpe SJ, Thorpe SJ (2005) Spike times make sense. Trends Neurosci 28:1–4.

Vreeswijk C v., Sompolinsky H (1996) Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. Science (80-) 274:1724–1726.

Wang C, Huang JY, Bardy C, FitzGibbon T, Dreher B (2010) Influence of "feedback" signals on spatial integration in receptive fields of cat area 17 neurons. Brain Res 1328:34–48.

Wang C, Waleszczyk WJ, Burke W, Dreher B (2000) Modulatory influence of feedback projections from area 21a on neuronal activities in striate cortex of the cat. Cereb Cortex 10:1217–1232.

Wang X-J (2010) Neurophysiological and computational principles of cortical rhythms in cognition. Physiol Rev 90:1195–1268.

Waters RS, Favorov O, Mori a., Asanuma H (1982) Pattern of projection and physiological properties of cortico-cortical connections from the posterior bank of the ansate sulcus to the motor cortex, area 4γ, in the cat. Exp Brain Res 48:335–344.

Watson C, Kirkaldie M, Paxinos G (2010) The Brain: An Introduction to Functional Neuroanatomy.

Wehr M, Zador AM (2003) Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature 426:442–446.

Werner G, Mountcastle V (1965) Neural activity in mechanoreceptive cutaneous afferents: stimulus-response relations, Weber functions, and information transmission. J Neurophysiol.

Wouterlood FG, Mugnaini E, Osen KK, Dahl a L (1984) Stellate neurons in rat dorsal cochlear nucleus studies with combined Golgi impregnation and electron microscopy: synaptic connections and mutual coupling by gap junctions. J Neurocytol 13:639–664.

Yordanova J, Kolev V, Kirov R (2012) Brain oscillations and predictive processing. Front Psychol 3:1–2.

York GK, Steinberg D a (2006) An introduction to the life and work of John Hughlings Jackson with a catalogue raisonné of his writings. Med Hist Suppl:3–157.

York GK, Steinberg D a. (2011) Hughlings Jackson's neurological ideas. Brain 134:3106–3113.

Yu Y, Shu Y, McCormick D a (2008) Cortical action potential backpropagation explains spike threshold variability and rapid-onset kinetics. J Neurosci 28:7260–7272.

Zarzecki P, Blum P, Bakker D, Herman D (1983) Convergence of sensory inputs upon projection neurons of somatosensory cortex: Vestibular, neck, head, and forelimb inputs. Exp Brain Res 50:408.

Zhang K, Sejnowski TJ (2000) A universal scaling law between gray matter and white matter of cerebral cortex. Proc Natl Acad Sci U S A 97:5621–5626.

Zhang LI, Tan AYY, Schreiner CE, Merzenich MM (2003) Topography and synaptic shaping of direction selectivity in primary auditory cortex. Nature 424:201–205.

Zhang S, Xu M, Kamigaki T, Hoang Do JP, Chang WC, Jenvay S, Miyamichi K, Luo L, Dan Y (2014) Long-range and local circuits for top-down modulation of visual cortex processing. Science (80-) 345:660–665.

Appendix

Table: references for the electrophysiological data

	Title	Year	Author
1	Prolonged synaptic integration in perirhinal cortical neurons.	2000	Brown TH
2	Subtype-specific dendritic Ca(2+) dynamics of inhibitory interneurons in the rat visual cortex.	2010	Rhie DJ
3	The roles of somatostatin-expressing (GIN) and fast-spiking inhibitory interneurons in UP-DOWN states of mouse neocortex.	2010	Connors BW
4	Glutamatergic nonpyramidal neurons from neocortical layer VI and their comparison with pyramidal and spiny stellate neurons.	2009	Lambolez B
5	Maturation of intrinsic and synaptic properties of layer 2/3 pyramidal neurons in mouse auditory cortex.	2008	Reyes AD
6	Characterization of neuronal migration disorders in neocortical structures. II. Intracellular in vitro recordings.	1998	Zilles K
7	Epileptogenesis following neocortical trauma from two sources of disinhibition.	1997	Benardo LS
8	Specialized cortical subnetworks differentially connect frontal cortex to parahippocampal areas.	2012	Kawaguchi Y
9	Sensory experience alters cortical connectivity and synaptic function site specifically.	2007	Finnerty GT
10	Background synaptic activity is sparse in neocortex.	2006	Helmchen F
11	Mechanisms and consequences of action potential burst firing in rat neocortical pyramidal neurons.	1999	Stuart GJ

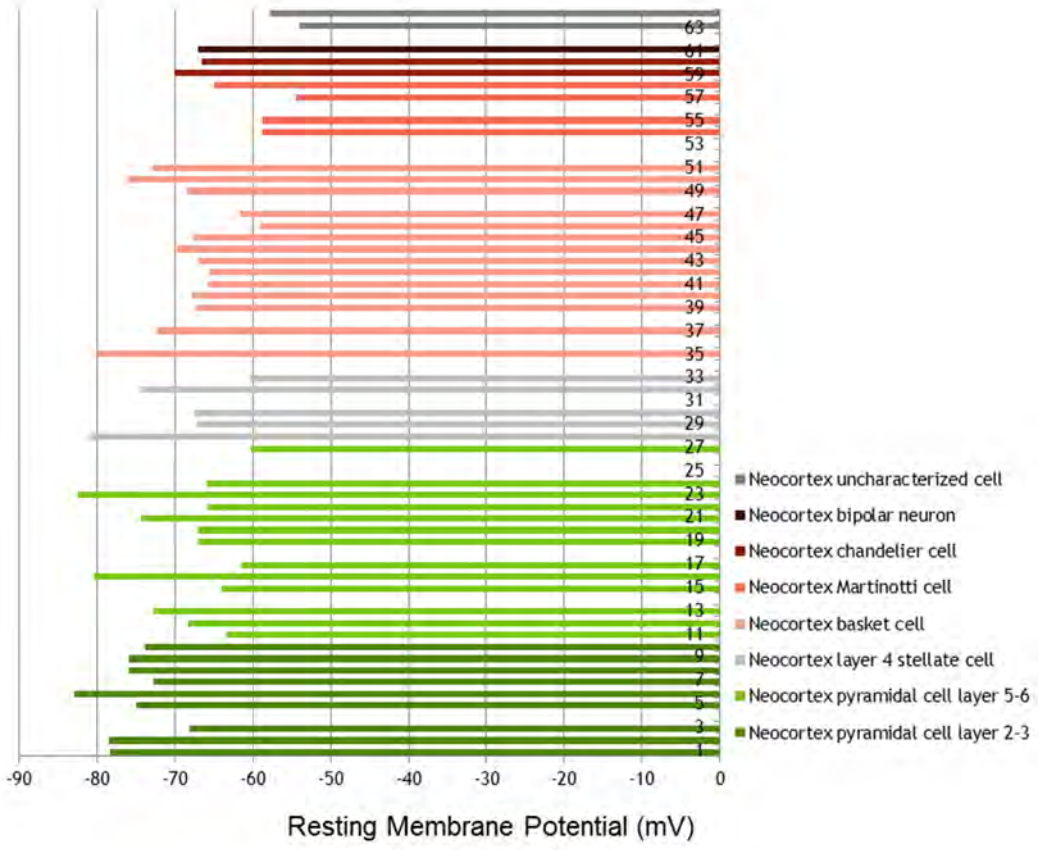
12	Properties of layer 6 pyramidal neuron apical dendrites.	2010	Larkum ME
13	Early exposure to alcohol leads to permanent impairment of dendritic excitability in neocortical pyramidal neurons.	2012	Larkum ME
14	GABAA receptor-mediated currents in interneurons and pyramidal cells of rat visual cortex.	1998	Prince DA
15	Differential effects of Na+-K+ ATPase blockade on cortical layer V neurons.	2010	Prince DA
16	Action potential initiation and propagation in layer 5 pyramidal neurons of the rat prefrontal cortex: absence of dopamine modulation.	2003	Stuart GJ
17	GABAergic synaptic inhibition is reduced before seizure onset in a genetic model of cortical malformation.	2006	Lee KS
18	Glutamatergic nonpyramidal neurons from neocortical layer VI and their comparison with pyramidal and spiny stellate neurons.	2009	Lambolez B
19	Morphological and physiological characterization of layer VI corticofugal neurons of mouse primary visual cortex.	2003	Yuste R
20	Epileptogenesis following neocortical trauma from two sources of disinhibition.	1997	Benardo LS
21	Increased excitability and inward rectification in layer V cortical pyramidal neurons in the epileptic mutant mouse Stargazer.	1997	Noebels JL
22	Specialized cortical subnetworks differentially connect frontal cortex to parahippocampal areas.	2012	Kawaguchi Y
23	Enhanced function of prefrontal serotonin 5-HT(2) receptors in a rat model of psychiatric vulnerability.	2010	Vaidya VA
24	Dopamine and corticotropin-releasing factor synergistically alter basolateral amygdala-to-medial prefrontal cortex synaptic transmission: functional switch after chronic cocaine administration.	2008	Gallagher JP

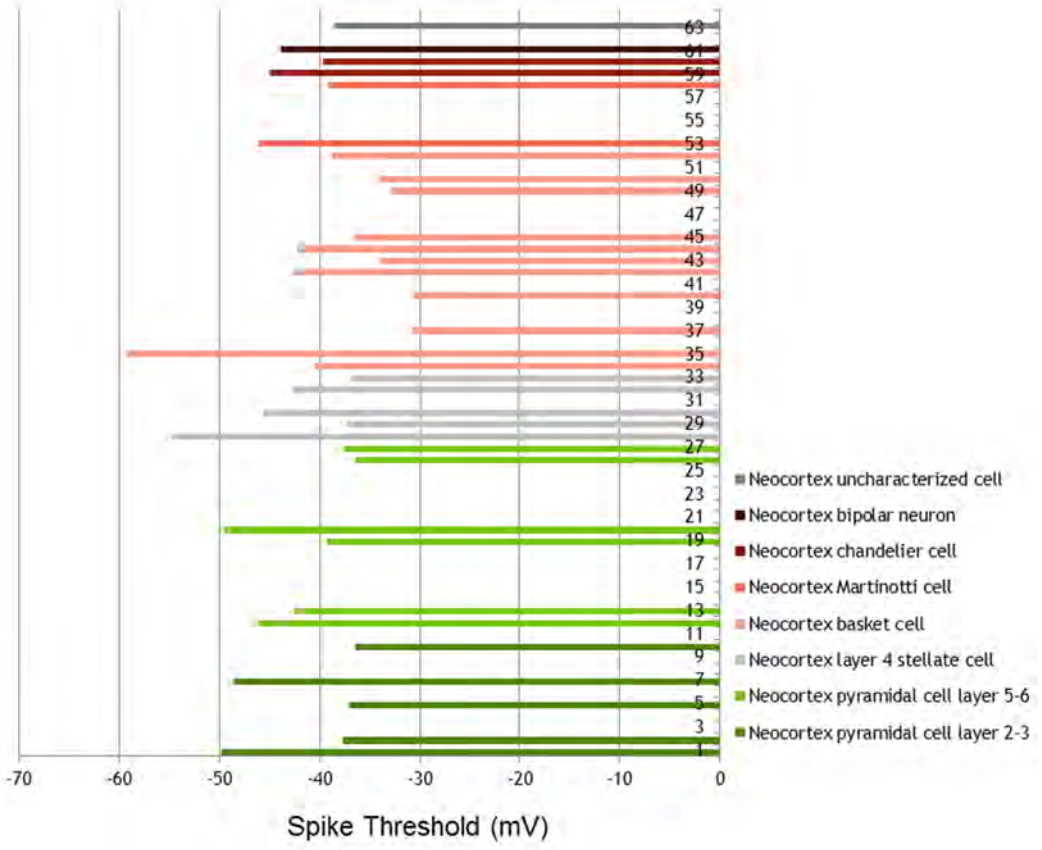
25	Postnatal development of synaptic transmission in local networks of L5A pyramidal neurons in rat somatosensory cortex.	2007	Sakmann B
26	Flexible spike timing of layer 5 neurons during dynamic beta oscillation shifts in rat prefrontal cortex.	2009	Mansvelder HD
27	Electrophysiological Abnormalities in Both Axotomized and Nonaxotomized Pyramidal Neurons following Mild Traumatic Brain Injury.	2012	Jacobs KM
28	Predomice of late-spiking neurons in layer VI of rat perirhinal cortex.	2001	Brown TH
29	Characterization of thalamocortical responses of regular-spiking and fast-spiking neurons of the mouse auditory cortex in vitro and in silico.	2012	Reyes AD
30	Response sensitivity of barrel neuron subpopulations to simulated thalamic input.	2010	Pinto DJ
31	Glutamatergic nonpyramidal neurons from neocortical layer VI and their comparison with pyramidal and spiny stellate neurons.	2009	Lambolez B
32	Auditory thalamocortical transmission is reliable and temporally precise.	2005	Metherate R
33	mGluR5 in cortical excitatory neurons exerts both cell-autonomous and -nonautonomous influences on cortical somatosensory circuit formation.	2010	Lu HC
34	Mechanisms of dopamine activation of fast-spiking interneurons that exert inhibition in rat prefrontal cortex.	2002	Yang CR
35	Predomice of late-spiking neurons in layer VI of rat perirhinal cortex.	2001	Brown TH
36	Major differences in inhibitory synaptic transmission onto two neocortical interneuron subclasses.	2003	Prince DA
37	Subtype-specific dendritic Ca(2+) dynamics of inhibitory interneurons in the rat visual cortex.	2010	Rhie DJ
38	GABAA receptor-mediated currents in interneurons and pyramidal cells of rat visual cortex.	1998	Prince DA

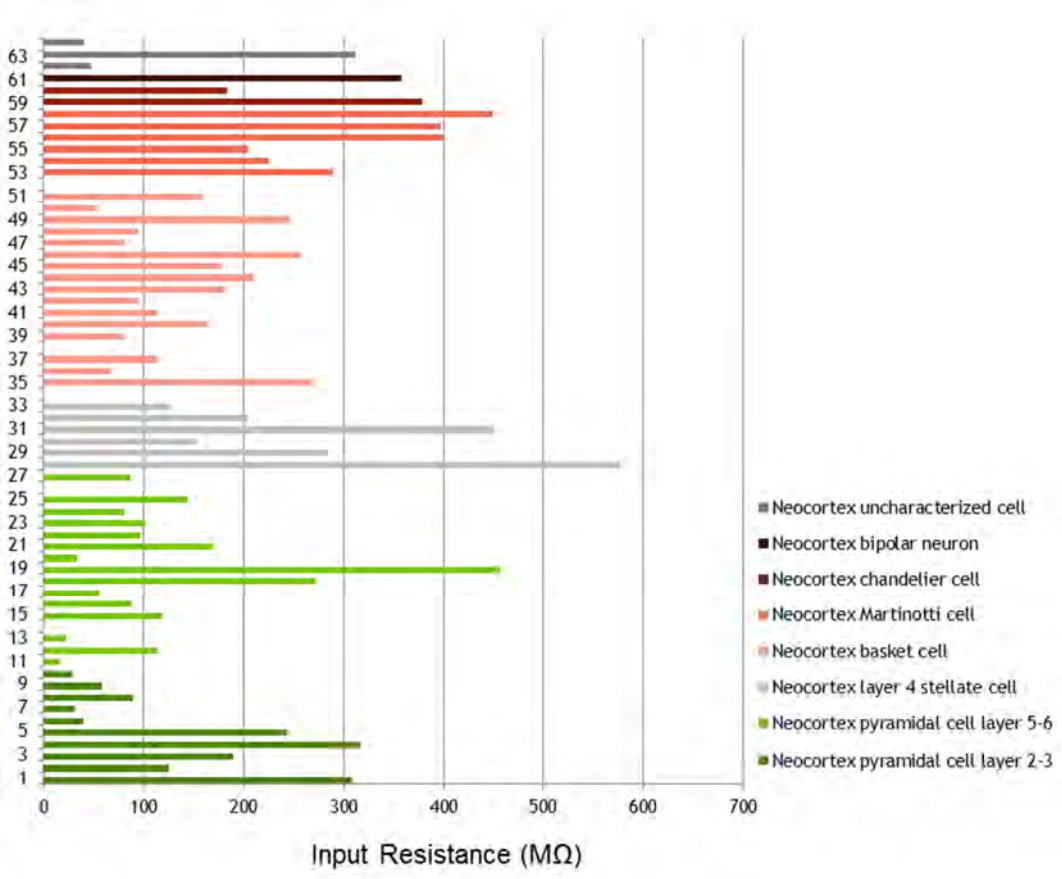
39	Differential effects of Na+-K+ ATPase blockade on cortical layer V neurons.	2010	Prince DA
40	Characterization of thalamocortical responses of regular-spiking and fast-spiking neurons of the mouse auditory cortex in vitro and in silico.	2012	Reyes AD
41	The roles of somatostatin-expressing (GIN) and fast-spiking inhibitory interneurons in UP-DOWN states of mouse neocortex.	2010	Connors BW
42	Response sensitivity of barrel neuron subpopulations to simulated thalamic input.	2010	Pinto DJ
43	Parvalbumin-positive basket interneurons in monkey and rat prefrontal cortex.	2008	Krimer LS
44	Auditory thalamocortical transmission is reliable and temporally precise.	2005	Metherate R
45	Functional properties of fast spiking interneurons and their synaptic connections with pyramidal cells in primate dorsolateral prefrontal cortex.	2005	Lewis DA
46	Neuregulin-1 signals from the periphery regulate AMPA receptor sensitivity and expression in GABAergic interneurons in developing neocortex.	2011	Nawa H
47	The synaptic representation of sound source location in auditory cortex.	2009	Margrie TW
48	Spike-timing-dependent plasticity of neocortical excitatory synapses on inhibitory interneurons depends on target cell type.	2007	Zhang XH
49	Physiologically distinct temporal cohorts of cortical interneurons arise from telencephalic Olig2-expressing precursors.	2007	Fishell G
50	Background synaptic activity is sparse in neocortex.	2006	Helmchen F
51	Developmental synaptic changes increase the range of integrative capabilities of an identified excitatory neocortical connection.	1999	Audinat E
52	Flexible spike timing of layer 5 neurons during dynamic beta oscillation shifts in rat prefrontal cortex.	2009	Mansvelder HD

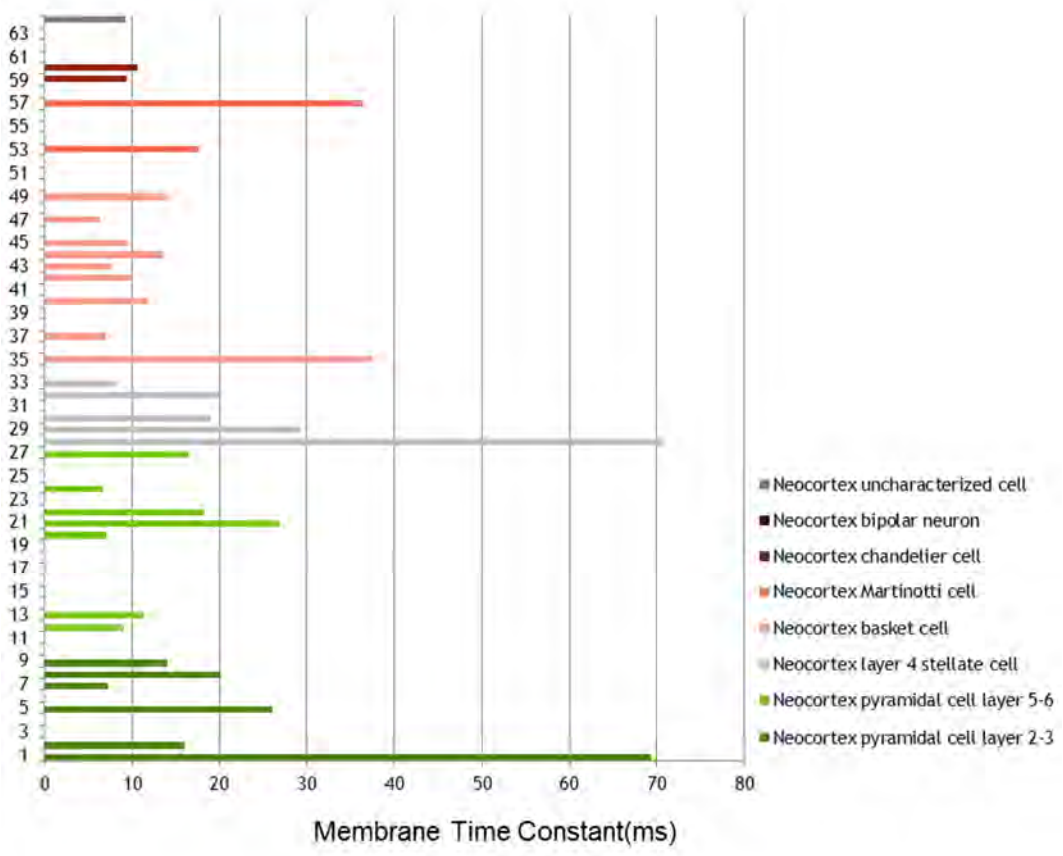
53	Anatomical physiological and molecular properties of Martinotti cells in the somatosensory cortex of the juvenile rat.	2004	Markram H
54	The roles of somatostatin-expressing (GIN) and fast-spiking inhibitory interneurons in UP-DOWN states of mouse neocortex.	2010	Connors BW
55	Electrophysiological classification of somatostatin-positive interneurons in mouse sensorimotor cortex.	2006	Prince DA
56	Spike-timing-dependent plasticity of neocortical excitatory synapses on inhibitory interneurons depends on target cell type.	2007	Zhang XH
57	Impaired inhibitory control of cortical synchronization in fragile X syndrome.	2011	Huntsman MM
58	Dense inhibitory connectivity in neocortex.	2011	Yuste R
59	Cluster analysis-based physiological classification and morphological properties of inhibitory neurons in layers 2-3 of monkey dorsolateral prefrontal cortex.	2005	Lewis DA
60	Functional properties of fast spiking interneurons and their synaptic connections with pyramidal cells in primate dorsolateral prefrontal cortex.	2005	Lewis DA
61	Functional characterization of intrinsic cholinergic interneurons in the cortex.	2007	Monyer H
62	Specificity in the interaction of HVA Ca2+ channel types with Ca2+-dependent AHPs and firing behavior in neocortical pyramidal neurons.	1998	Foehring RC
63	Fear conditioning and extinction differentially modify the intrinsic excitability of infralimbic neurons.	2008	Porter JT
64	The synaptic representation of sound source location in auditory cortex.	2009	Margrie TW

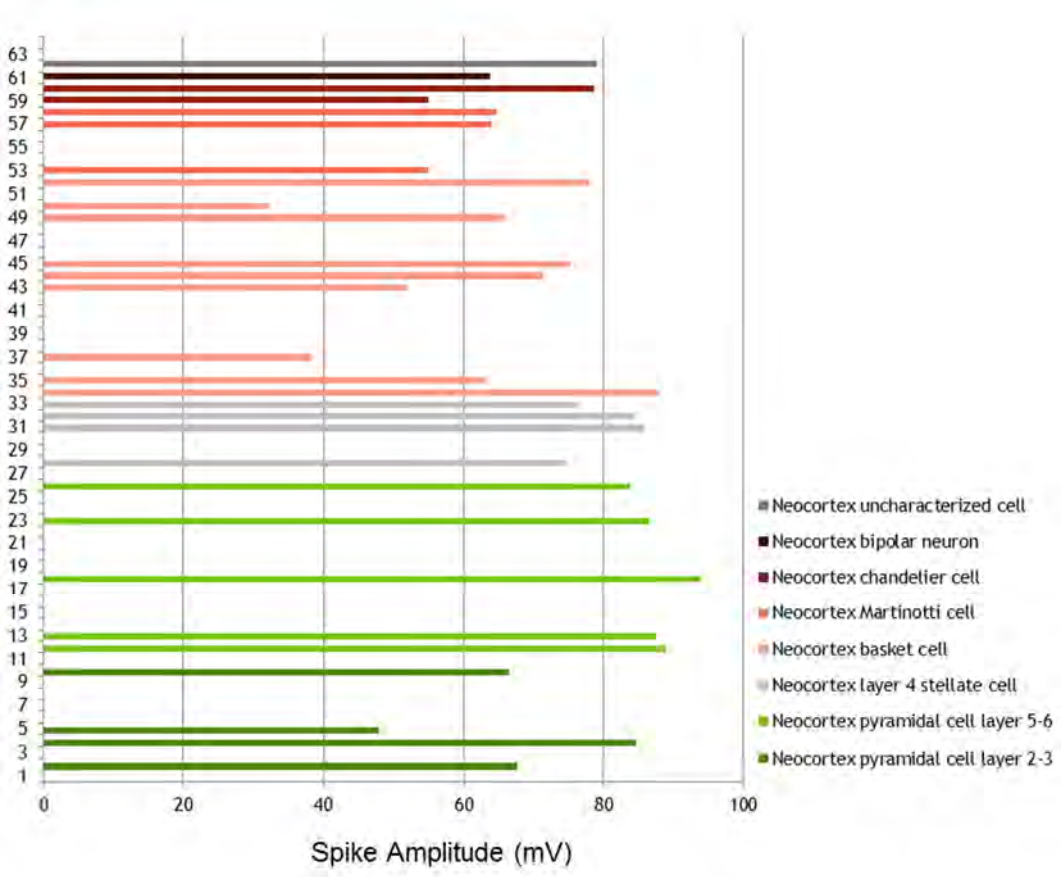
* Data obtained from NeuroElectro.org

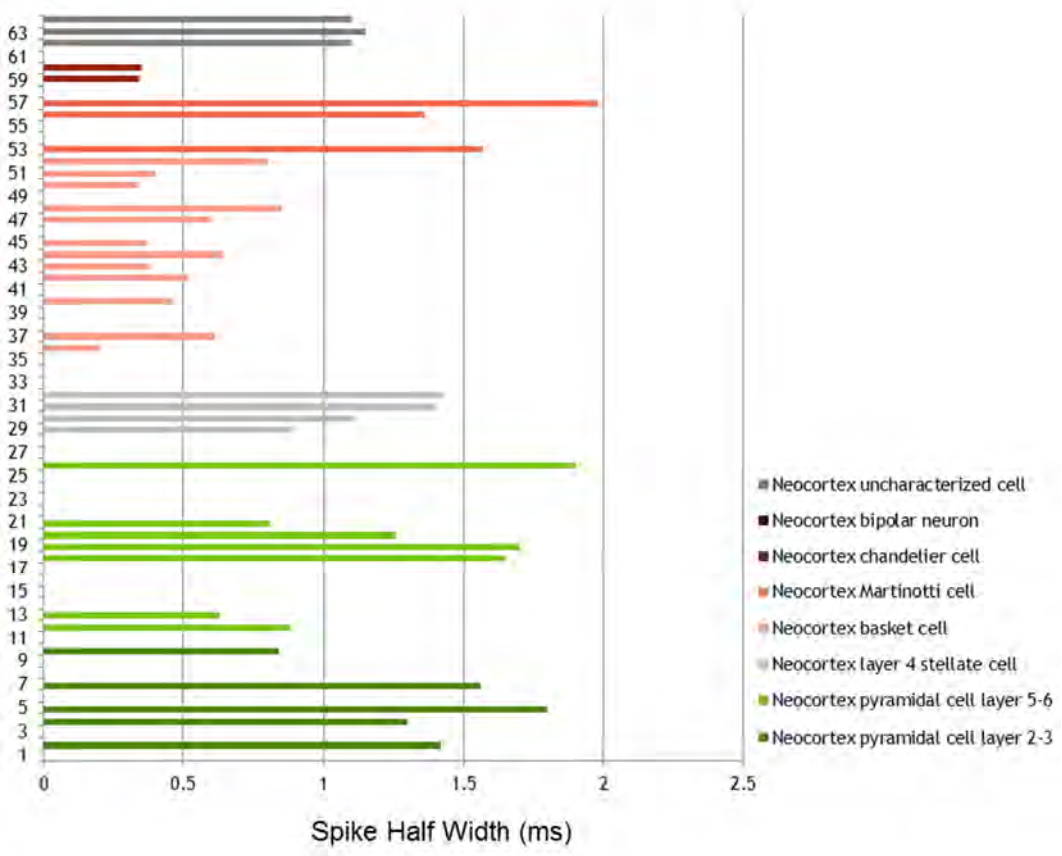

63
61
59
57
55
53
51
49
47
45
43
41
39
37
35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
-90
-80
-70
-60
-50
-40
-30
-20
-10
0
Resting Membrane Potential (mV)
Neocortex uncharacterized cell
Neocortex bipolar neuron
Neocortex chandelier cell
Neocortex Martinotti cell
Neocortex basket cell
Neocortex layer 4 stellate cell
Neocortex pyramidal cell layer 5-6
Neocortex pyramidal cell layer 2-3

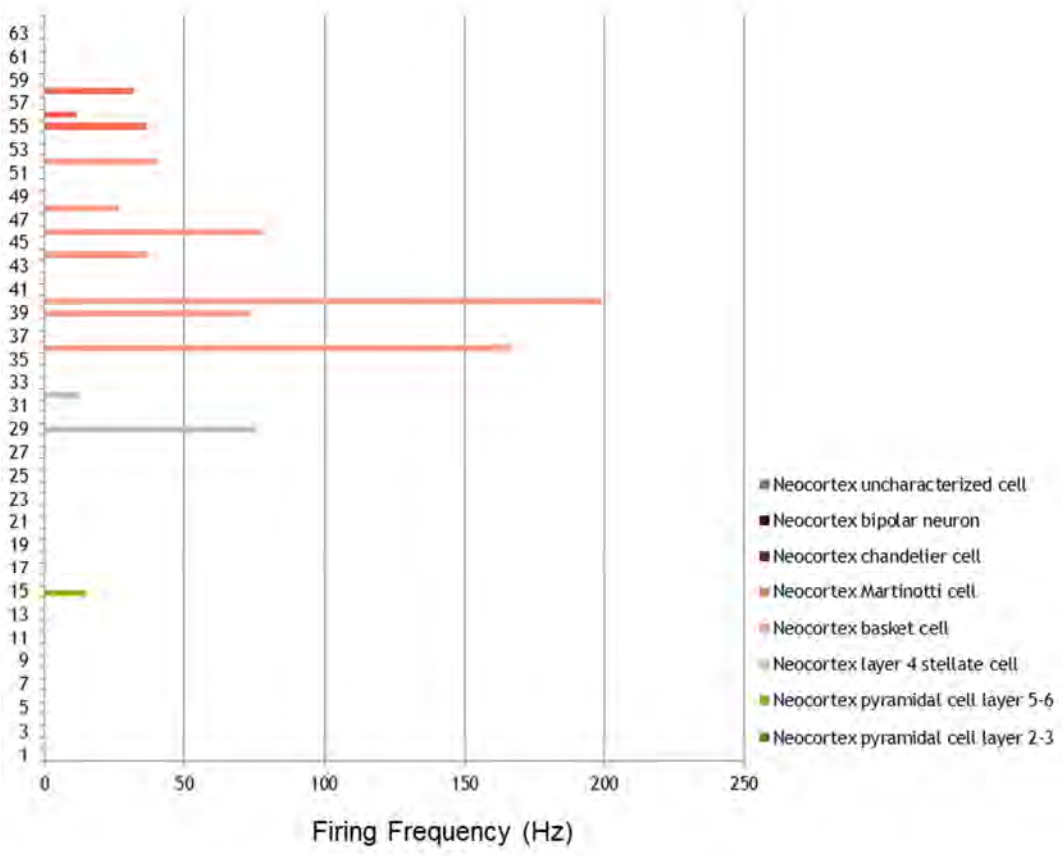

63
61
59
57
55
53
51
49
47
45
43
41
39
37
35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
-70
-60
-50
-40
-30
-20
-10
0
Spike Threshold (mV)
Neocortex uncharacterized cell
Neocortex bipolar neuron
Neocortex chandelier cell
Neocortex Martinotti cell
Neocortex basket cell
Neocortex layer 4 stellate cell
Neocortex pyramidal cell layer 5-6
Neocortex pyramidal cell layer 2-3

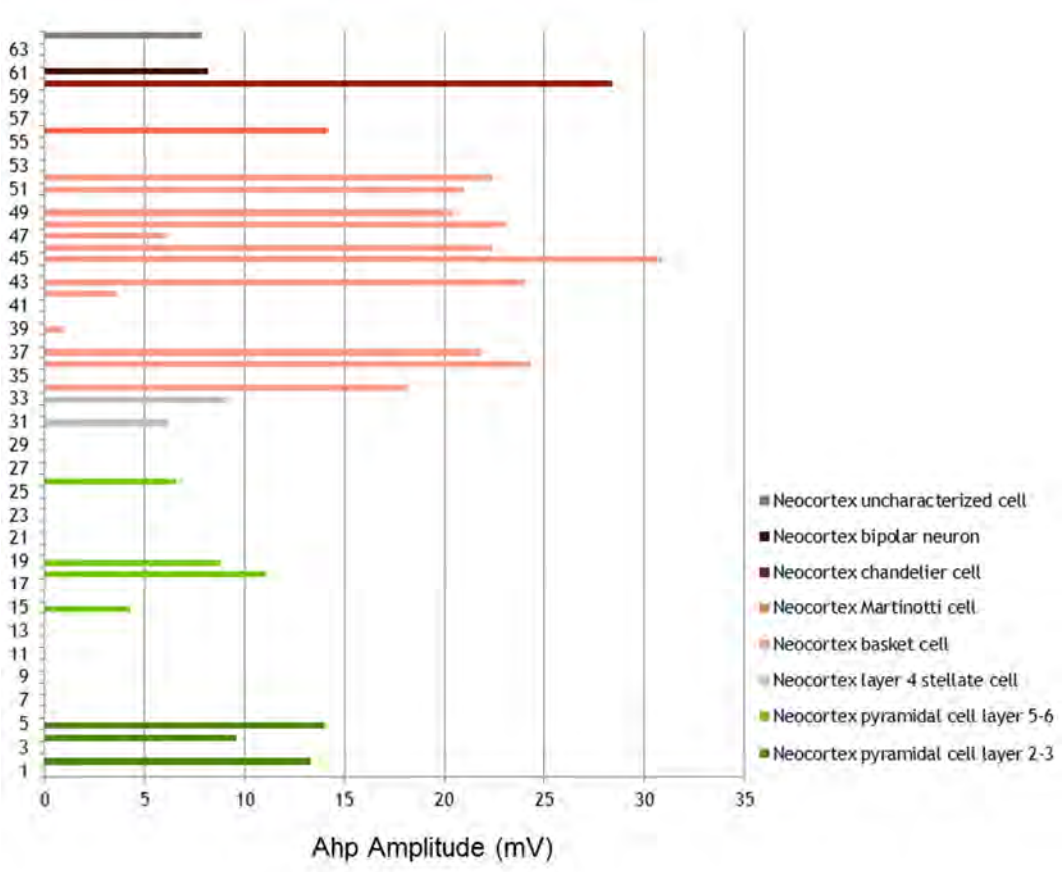

63
61
59
57
55
53
51
49
47
45
43
41
39
37
35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
0
100
200
300
400
500
600
700
Input Resistance (MΩ)
Neocortex uncharacterized cell
Neocortex bipolar neuron
Neocortex chandelier cell
Neocortex Martinotti cell
Neocortex basket cell
Neocortex layer 4 stellate cell
Neocortex pyramidal cell layer 5-6
Neocortex pyramidal cell layer 2-3

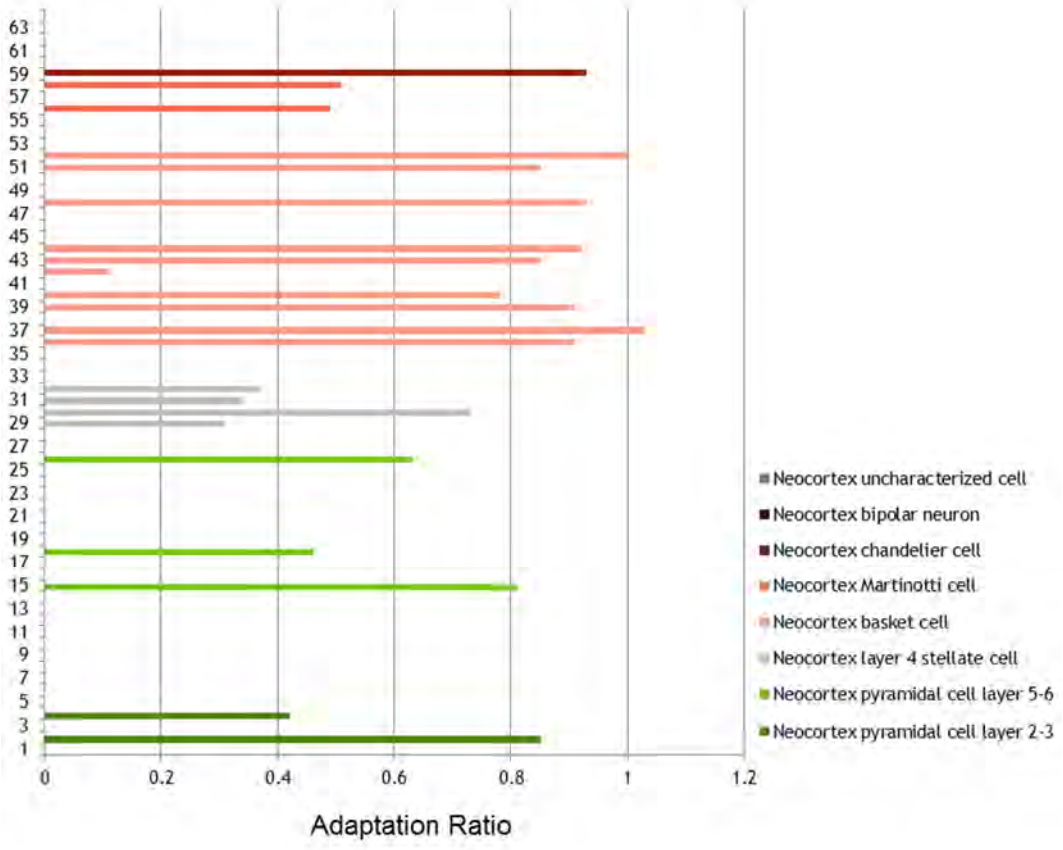

63
61
59
57
55
53
51
49
47
45
43
41
39
37
35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
0
10
20
30
40
50
60
70
80
Membrane Time Constant(ms)
Neocortex uncharacterized cell
Neocortex bipolar neuron
Neocortex chandelier cell
Neocortex Martinotti cell
Neocortex basket cell
Neocortex layer 4 stellate cell
Neocortex pyramidal cell layer 5-6
Neocortex pyramidal cell layer 2-3

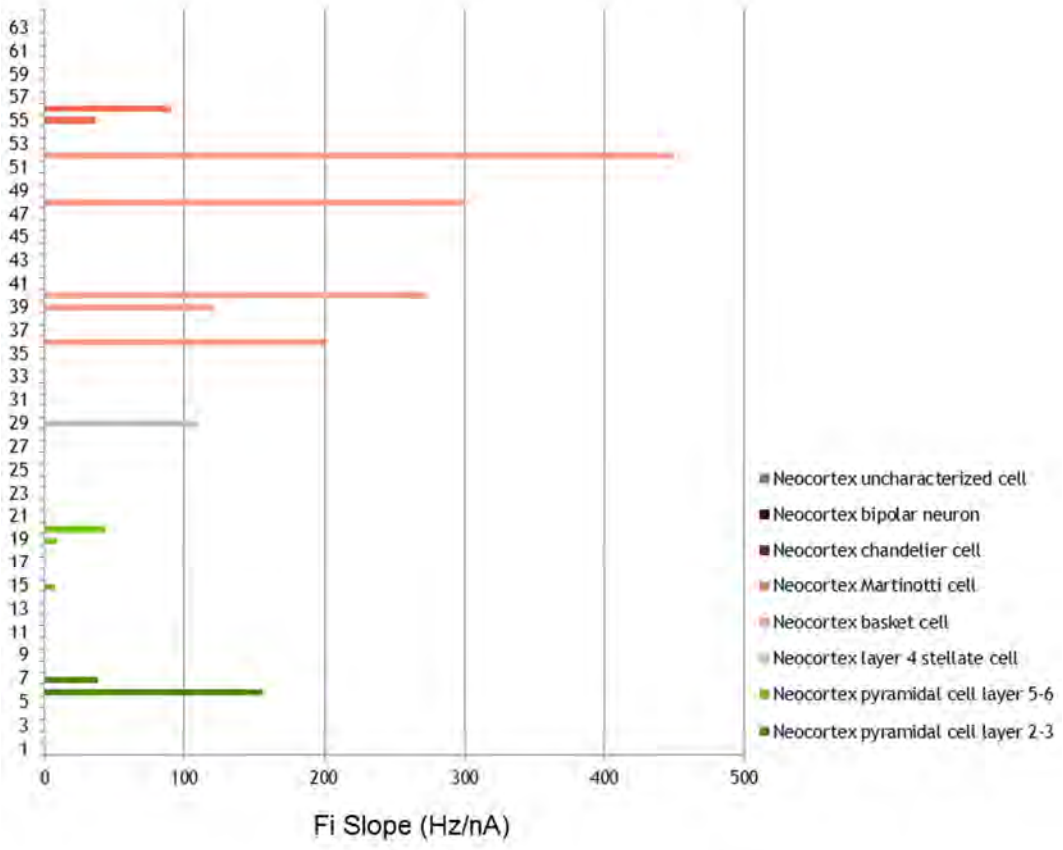

63
61
59
57
55
53
51
49
47
45
43
41
39
37
35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
0
20
40
60
80
100
Spike Amplitude (mV)
Neocortex uncharacterized cell
Neocortex bipolar neuron
Neocortex chandelier cell
Neocortex Martinotti cell
Neocortex basket cell
Neocortex layer 4 stellate cell
Neocortex pyramidal cell layer 5-6
Neocortex pyramidal cell layer 2-3


63
61
59
57
55
53
51
49
47
45
43
41
39
37
35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
0
0.5
1
1.5
2
2.5
Spike Half Width (ms)
Neocortex uncharacterized cell
Neocortex bipolar neuron
Neocortex chandelier cell
Neocortex Martinotti cell
Neocortex basket cell
Neocortex layer 4 stellate cell
Neocortex pyramidal cell layer 5-6
Neocortex pyramidal cell layer 2-3


63
61
59
57
55
53
51
49
47
45
43
41
39
37
35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
0
50
100
150
200
250
Firing Frequency (Hz)
Neocortex uncharacterized cell
Neocortex bipolar neuron
Neocortex chandelier cell
Neocortex Martinotti cell
Neocortex basket cell
Neocortex layer 4 stellate cell
Neocortex pyramidal cell layer 5-6
Neocortex pyramidal cell layer 2-3


63
61
59
57
55
53
51
49
47
45
43
41
39
37
35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
0
5
10
15
20
25
30
35
Ahp Amplitude (mV)
Neocortex uncharacterized cell
Neocortex bipolar neuron
Neocortex chandelier cell
Neocortex Martinotti cell
Neocortex basket cell
Neocortex layer 4 stellate cell
Neocortex pyramidal cell layer 5-6
Neocortex pyramidal cell layer 2-3


63
61
59
57
55
53
51
49
47
45
43
41
39
37
35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
0
0.2
0.4
0.6
0.8
1
1.2
Adaptation Ratio
Neocortex uncharacterized cell
Neocortex bipolar neuron
Neocortex chandelier cell
Neocortex Martinotti cell
Neocortex basket cell
Neocortex layer 4 stellate cell
Neocortex pyramidal cell layer 5-6
Neocortex pyramidal cell layer 2-3


63
61
59
57
55
53
51
49
47
45
43
41
39
37
35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
0
100
200
300
400
500
Fi Slope (Hz/nA)
Neocortex uncharacterized cell
Neocortex bipolar neuron
Neocortex chandelier cell
Neocortex Martinotti cell
Neocortex basket cell
Neocortex layer 4 stellate cell
Neocortex pyramidal cell layer 5-6
Neocortex pyramidal cell layer 2-3

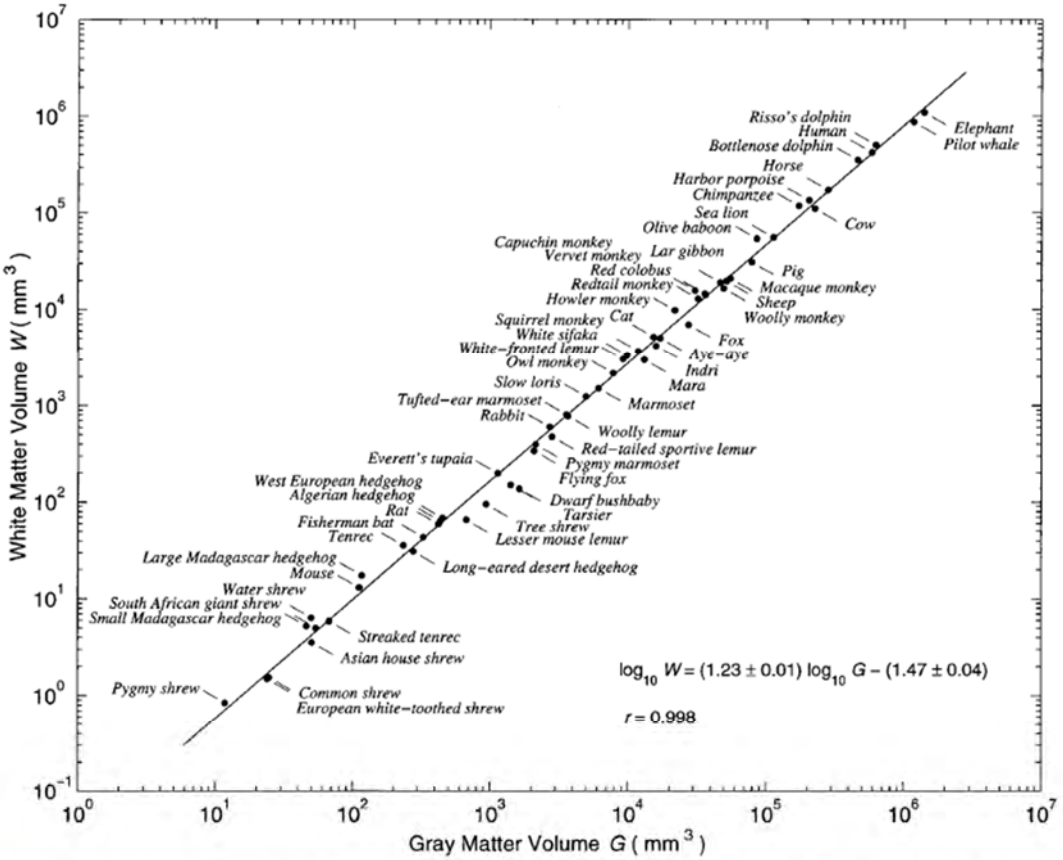


Fig. 2. Cortical white and gray matter volumes of various species (n = 59) are related by a power law that spans five to six orders of magnitude. Most data points are based on measurement of a single adult animal. The line is the least squares fit, with a slope around 1.23 ± 0.01 (mean ± SD). The average and median deviations of the white matter volumes from the regression line are, respectively, 18% and 13% on a linear scale. *Sources of data*: If the same species appeared in more than one source below, the one mentioned earlier was used. All 38 species in table 2 in ref. 3 were taken, including 23 primates, 2 tree shrews, and 13 insectivores. Another 11 species were taken from table 2 in ref. 8, including 3 primates, 2 carnivores, 4 ungulates, and 2 rodents. Five additional species came from table 1 in ref. 11, including 1 elephant and 4 cetaceans. The data point for the mouse (G = 112 mm^3 and W = 13 mm^3) was based on ref. 30, and that for the rat (G = 425 mm^3 and W = 59 mm^3) was measured from the serial sections in a stereotaxic atlas (42). The estimates for the fisherman bat (*Noctilio leporinus*, G = 329 mm^3 and W = 43 mm^3) and the flying fox (*Pteropus lylei*, G = 2,083 mm^3 and W = 341 mm^3) were based on refs. 43 and 44, with the ratios of white and gray matters estimated roughly from the section photographs in the papers. The sea lion data (*Zalophus californianus*, G = 113,200 mm^3 and W = 56,100 mm^3) were measured from the serial sections at the website given in the legend to Fig. 1, with shrinkage correction.